

Treading water

Some oceanographers were hoping that a commission chaired by James Watkins would propose more radical reform of ocean research, education and exploration.

This week's report from the US Commission on Ocean Policy (see page 787) should lend valuable impetus to proposed legislation that would strengthen the mechanisms and financial support for ocean research.

As James Watkins, the commission's chair, has repeatedly pointed out, its work represents a rare opportunity to bring important issues, such as the declining health of marine ecosystems and bad fisheries management, to the fore. The exploration and management of the oceans is a widely neglected government function, not just in the United States, and the first review of ocean policy in 35 years provides a rare opportunity to overhaul it.

But critics of the present set-up are already arguing that the commission may have missed its chance. They say that its central proposal — to create a National Ocean Council at the White House — is insufficient to reconcile the conflicting policies that dog ocean management.

The commission could instead have recommended taking the National Marine Fisheries Service (NMFS), which regulates fish and other marine life, along with its parent agency, the National Oceanic and Atmospheric Administration (NOAA), out of the US Department of Commerce. Scientists at the NMFS and the rest of NOAA often say that they are constantly looking over their shoulders at pressures from commercial interests, which should not be allowed to override the wider public interest in sustainable ocean management.

Critics say that the NMFS would work better inside the Depart-

ment of the Interior, or that NOAA could become an independent agency, like the space agency NASA, to whose budget and national profile oceanographers have often cast an envious eye. An alternative would be to create a cabinet-level Department of Natural Resources, combining agencies such as the NMFS with the interior department's resource services, as President Jimmy Carter proposed 25 years ago. This kind of radical surgery was endorsed last year by the non-governmental Pew Oceans Commission (see *Nature* **423**, 577; 2003). But Watkins' panel apparently felt that proposing a new agency in the current climate in Washington would be an exercise in futility.

Instead it urges government to "strengthen NOAA and improve the federal agency structure", and urges the doubling of the agency's research budget. It also suggests the creation of a trust fund, whereby the oil and gas industry, and other commercial users of offshore territory, would pay levies to help with ocean management.

Fortunately, the oceans have plenty of friends on both sides of the aisle in Congress — as does Admiral Watkins. Perhaps it is a shame that the commission didn't strike a bolder course. Even so, the report, along with the Pew commission's work and last year's National Academy of Sciences study on ocean exploration, puts Congress in a good position to legislate for the future health of the oceans. It should pass laws this year to strengthen NOAA and create a National Ocean Council at the White House, and ensure adequate support for a proposed integrated ocean-observing system. ■

Organic farming enters the mainstream

Conventional agriculture needs to plough furrows of knowledge that were dismissed until recently as marginal and esoteric.

Ten years ago, few supermarkets stocked organic food, and those that did carried little more than some wilted lettuce and a basket of spotty apples — nothing to suggest that organic agriculture had a sparkling future. But now the future is here, and whole aisles of organic products in many stores command higher prices than food grown by conventional, intensive methods.

Researchers into organic methods have enjoyed a corresponding upsurge of interest in their work. Once relegated to the fringes of agricultural colleges and research stations, people who study composting and soil-organism biodiversity, for example, are now being taken seriously (see pages 792–798). Even the most intensive farmers are conscious of the need for farming methods that will be sustainable in the long term, and the research emphasis has shifted from maximizing yield to finding sustainable techniques. It is a moment to be savoured by proponents of organic agriculture and soil conservation, and presents an opportunity that must not be squandered.

Public discussions of organic farming, particularly in the news media, tend to be sharply polarized. Advocates insist that conventional farming exhausts the land and is unsustainable. Critics question the wisdom of farming with manure, and suggest that switching to less efficient methods is unethical when so many people in the world are hungry.

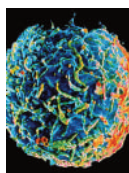
These debates tend to be heated in the scientific literature, too,

with representatives of each point of view finding studies that back them up. Evidence can be produced to show that organic farms have higher yields or lower yields, produce more water-fouling nitrogen runoff or less. The resultant morass makes it tough for farmers and policy-makers to pin down the relevant facts.

This doesn't mean that the studies are necessarily flawed. It simply reflects the complexity of the systems being studied, and the number of variables involved. Straight comparisons of organic and conventional practice can often be misleading. Organic and conventional growers generally use different strains, for example, emphasizing pest resistance and yield, respectively. And organic farms change year to year, with yields improving and runoff decreasing over time as organic matter builds up and weed seeds disappear.

A better approach is to focus studies on particular issues of interest, such as soil conservation, nutrient retention and pest control. These should be done by collaborations between researchers who may have held differing views in the past. For too long, agricultural studies have tended to involve like-minded researchers and to be supported by a non-profit group, industrial company or government department that has an interest in the outcome. Now that such issues as soil conservation are acknowledged to be vital to all farmers, researchers with differing perspectives on the organic issue need to share their insights and seek out approaches that work. ■





Blocking B cells
Cancer drug may join
arsenal to combat
autoimmune diseases
p786



Deep thoughts
US panel reports on
ways to revitalize
ocean research
p787



Iron awe
Sea fertilizer could
create carbon sinks to
fight global warming
p788



Fresh culprits
Foxes and shrews join
civets on the list of
SARS suspects
p790

Boston locals fight government scheme for bioterror defence lab

Rex Dalton

Plans to build a \$178-million infectious diseases laboratory at Boston University have triggered strong objections from local scientists and residents, threatening the US government's plans to shore up its defences against biological agents.

Around 150 scientists from the Boston area released a letter on 13 April calling on city officials and Boston University trustees to halt plans to build the National Biocontainment Laboratory — a biosafety level 4 facility that could handle the world's most dangerous pathogens — in the city's South End.

They say that they object to the facility on scientific, safety and political grounds, and have joined residents' groups who claim that the laboratory is being foisted on one of Boston's poorest neighbourhoods, next door to Boston University Medical Center. The neighbourhood, once the most famous Irish ghetto in the United States, is now populated mainly by African Americans.

Boston City Council was due to hold a public hearing on 20 April on the safety of the proposed laboratory. But three of the council's 13 members have already declared



David Ozonoff: retracted support for biodefence lab, saying scientists were being misled.

their support for a proposed municipal law that would outlaw the facility. Last week, anxious university administrators were trying to rally public support for the project, and its supporters plan to run newspaper advertisements in its defence.

David Ozonoff, a physician at Boston University and former chair of its environmental

health department, says he was originally in favour of the project, but had turned against it because administrators were "misleading" the scientific community. Military and law-enforcement agencies would be able to "stipulate whatever they want" from the facility, he claims, adding that the lab's existence could increase the likelihood that dangerous pathogens would end up in terrorists' hands.

Mark Klempner, a physician, associate provost of the university and principal investigator on the laboratory project, declined to discuss the objections. But a spokeswoman for the university said that it was dedicated to creating a dynamic and safe facility.

The Boston laboratory is one of two mainstays of the National Institutes of Health's response to the threat of bioterrorism; a similar containment laboratory is also to be built at the University of Texas in Galveston. A federal public hearing in Galveston on 31 March attracted only a single dissenting voice.

Awards to construct the two facilities were made last September by the National Institute of Allergy and Infectious Diseases after an intense competition between possible host universities. Nine biosafety level 3 laboratories were also announced by the agency.

Officials at the National Institutes of Health say they are confident that the laboratory will still be built; university officials have hired architects for the project.

Sheldon Krinsky, a professor of urban planning at Tufts University in Medford, Massachusetts, and an opponent of the project, says it is too early to tell if the Boston site is doomed. "This is a top-down thing, presented by the university to the community," he says.

Objectors complain that the lab should not be placed in a densely populated area where so many residents lack basic health insurance. "Scientists should be ashamed that they are taking advantage of poor people," says Daniel Goodenough, a cell biologist at Harvard Medical School.

But the university argues that the laboratory will bring jobs and tens of millions of dollars in research funding to a hard-pressed corner of the city.

Europe overpays research expenses

Alison Abbott, Munich

Over a fifth of the expenses claimed by researchers funded under the European Commission's Fifth Framework Programme of Research (FP5) should not have been paid, according to the European Court of Auditors.

The court studied 28 representative projects from the €13.7-billion (US\$16.3-billion), five-year programme, which ended in early 2003. Its report, published on 20 April, found a 21% error margin in the payouts.

Typical cases include the purchase of scientific equipment that was not covered by the terms of the grant, although the auditors say that the mistakes are due to mismanagement rather than fraud. More careful screening of the expenses could have

avoided the problem, the court says.

The commission worked closely with the court during the audit, which began more than a year before FP5 ended. Realizing the problem, it inserted a requirement for external auditing of grantees' eligible costs into its follow-up Sixth Framework Programme.

But this extra rule is a burden for scientists, concedes commission spokesman Fabio Fabbi. "The commission is criticized for loopholes that allow excessive spending, but also for being overly bureaucratic," he says, "We have to strike a balance." The FP5 rules were already "an unnecessary complication for participants", according to the auditors.

Health department lays down the law on scientific misconduct

Meredith Wadman, Washington

The US health department has issued comprehensive rules for handling scientific misconduct cases. And instead of kicking up a stink, most research organizations say that they find the regulations to their liking.

On 19 April, the Department of Health and Human Services issued a 27-page notice in the Federal Register, which publishes rules for US federal organizations. The department has narrowed its current definition of research misconduct and set out procedures for universities and government officials investigating allegations of scientific wrong-doing by federally funded researchers.

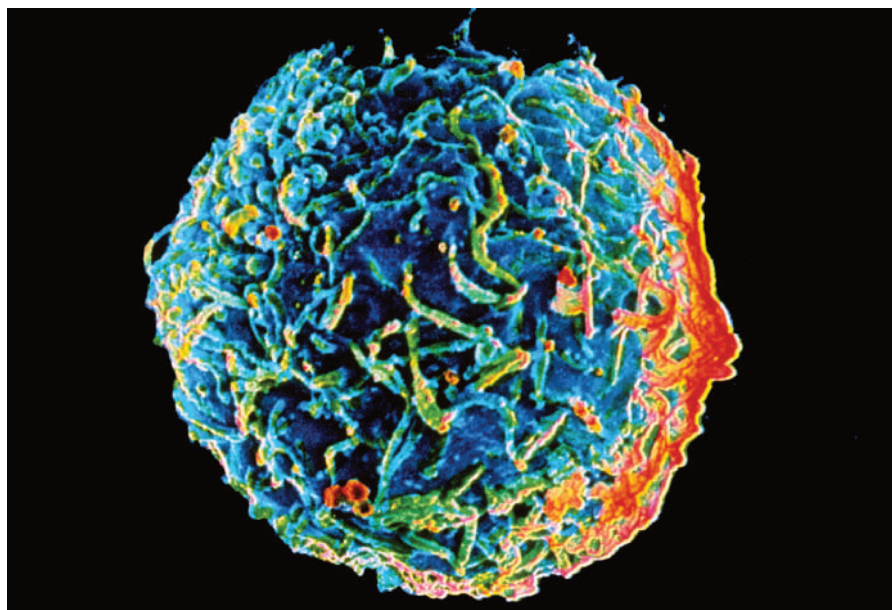
The regulations are probably the final chapter in a decade-long debate over what exactly constitutes scientific misconduct. And the department, which oversees most US biomedical research, has restricted the definition to “fabrication, falsification and plagiarism” — the offences that most researchers agree are egregious. The department has ditched the inclusion, in its current rules, of the ominous phrase “or other practices”, which scientists criticized as vague and overly broad.

Last week's notice expands the definition of misconduct to say that it can occur during the review of grant proposals and journal articles. It also limits the role of whistle-blowers in investigations, stating that complainants may not participate except as witnesses.

“What they have created is a consistent, clean definition,” says David Korn, senior vice-president for biomedical and health sciences research at the Association of American Medical Colleges. “Everybody understands what it means.”

Not everyone shares Korn's cheerful perspective. “They are now saying a whistle-blower is merely someone who reports, and then stands aside,” complains Vera Sharav, president of the Alliance for Human Research Protection, a New York-based patient advocacy group. “That is outrageous. Who else knows how to guide the investigation to evidence of the violations?”

Among other things, the revised policy requires complaints of misconduct to be filed within six years of the alleged event. The notice is open to public comment until 15 June, after which a final version will be published. ■



Can cancer drugs that attack B cells (above) also help to treat conditions such as rheumatoid arthritis?

Mouse opens door for study of autoimmune diseases

Erika Check, Washington

Biotechnology firm Genentech is pinning its hopes on a mouse model, in a bid to broaden the use of a billion-dollar cancer drug.

The company, based in South San Francisco, makes the drug Rituxan, which is used to treat non-Hodgkin's lymphoma. But clinical trials have suggested that the drug may also alleviate symptoms of autoimmune diseases, in which the immune system attacks the body's own cells. To test this idea, the company has created a mouse model that it says could speed up research on such diseases.

Rituxan works by targeting a specific protein carried on the surface of B cells, which form part of the body's immune response. By latching on to these cells it marks them for destruction, in the process clearing cancerous cells from the body. Genentech's new mouse model expresses the human B-cell protein recognized by Rituxan.

The model was announced on 19 April by Andrew Chan, Genentech's vice-president for immunology research, at the Federation of American Societies for Experimental Biology meeting in Washington DC. Scientists say that it could help fill a need for research on whether ‘depletion therapies’, such as Rituxan, can safely help patients with autoimmune diseases, by lending insight into the drugs' action. So far, many such therapies have yielded disappointing results in the clinic, but researchers hope that the new model could help to break that pattern.

“The preclinical models for depletion were really centred around cancer,” says Edward Clark, an immunologist at the

University of Washington in Seattle. “We need good preclinical models to study these treatments in autoimmune diseases as well.”

Chan told the meeting that the model has already been used to investigate key safety issues. For instance, clinical trials suggest that Rituxan eases symptoms of the autoimmune disease rheumatoid arthritis, in which the body destroys its own joints. But scientists had been puzzled by the fact that Rituxan does not seem to cause serious infections in the arthritis sufferers, even though it depletes their immune system of B cells.

Genentech scientists found that in mice some types of B cells survived the Rituxan treatment. Chan and his co-workers speculate that the drug kills only certain classes of B cell, leaving others behind to fight off disease. “Maybe this is why the arthritis patients keep fairly intact immune systems,” Chan says.

The finding might also have implications for how long and how well the drug works. Genentech is investigating such issues in a large clinical trial of Rituxan in rheumatoid arthritis patients. It is also testing Rituxan in patients with another autoimmune disease, called systemic lupus erythematosus.

Like many autoimmune diseases, lupus is poorly understood and difficult to treat. Clark, who has worked as a paid consultant to Genentech in the past, says that the new model could help answer questions that have held back this research, such as why depletion therapies only work in some autoimmune diseases, and whether the drugs can be made to work better. ■

Panel seeks fresh course for ocean research

Virginia Gewin

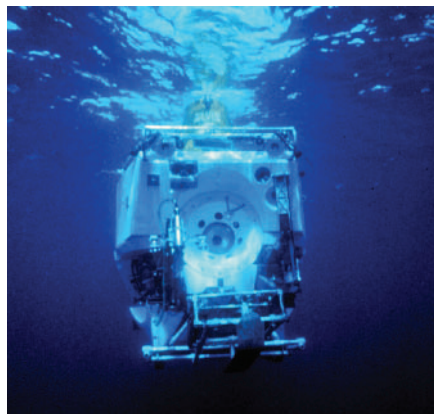
Ocean scientists in the United States have been waiting three years for a study, commissioned by President Bush, that they fervently hoped would revitalize the institutions that support their work. This week they got their report — but it remains unclear whether the revitalization will take place.

The US Commission on Ocean Policy, chaired by James Watkins — who served as energy secretary under Bush's father — recommended 200 reforms in the document it was expected to release on 20 April. The report calls for the establishment of a National Ocean Council at the White House, to coordinate the \$7 billion that the government spends each year on ocean management, research and education.

But the commission rejects calls for an independent ocean research agency to replace the National Oceanic and Atmospheric Administration (NOAA), which is part of the Department of Commerce. They say privately that strengthening NOAA is more feasible politically.

The panel calls for a doubling of ocean research funding over five years and for greater focus on ecosystem approaches to ocean management that address relationships between land, sea and air. It also says that regional councils should be set up to help manage fisheries, and that \$1.7 billion should be invested in an integrated ocean observing system, comprising satellites, buoys, sensors and computer systems, to be run by NOAA.

The commission's mandate was broader



Making a splash? An expert panel has called for a revamp of US oceanography.

than that of last year's independent Pew Oceans Commission (see *Nature* 423, 577; 2003), which focused only on living resources, but many of the recommendations are similar. However, the Pew panel backed the idea of setting up an independent agency to supplant NOAA.

"If the administration doesn't capitalize on both commission reports, it will be a real disservice to the nation," says Jane Lubchenco, a marine biologist at Oregon State University and member of the Pew panel.

Some fear that the Watkins report, which took longer than expected to appear, may have arrived too late to influence policy under the Bush administration. Nonetheless, advocates of ocean research are pursuing legislation in both the House and the Senate to set up the National Ocean Council

and implement other aspects of the report.

James Connaughton, chairman of the White House Council on Environmental Quality, says the Watkin commission's work has already spurred the administration to action. He points out that it recently sent officials to an international summit aimed at implementing an integrated global observation system, and sent legislation to Congress calling for more effective, market-based management tools for fisheries.

Ocean research supporters in the Senate have already passed bills on coastal development, the ocean observing system and research on the relationship between oceans and human health, and have introduced an ocean-exploration bill. And both houses will hold hearings on the more comprehensive Big Ocean Bill, known affectionately as Bob, before the end of the month.

But none of this activity will necessarily give ocean explorers the kind of public prestige to which they aspire. "We're spending billions to find out if there was water on Mars when we know there's water on Earth but don't know what shape it's in," says Jim Greenwood (Republican, Pennsylvania), co-chairman of the House Oceans Caucus.

But Watkins is optimistic that his panel's report will strengthen US ocean research in the long term. New approaches and investment are urgently needed, he says. Watkins doesn't think that other priorities in Washington — such as the war on terrorism — will bury his report. "We're a big nation," he says. "We ought to be able to walk and chew gum at the same time."

OAR/NURPWoods HOLE OCEANOGRAPHIC INST.

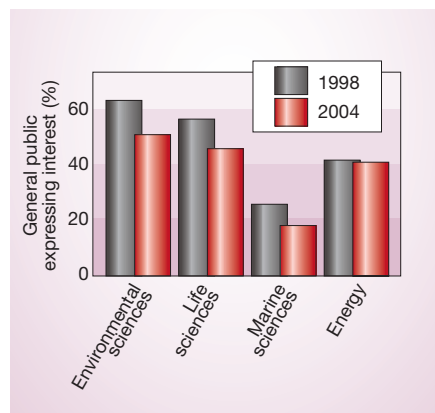
Japanese Nobels fail to inspire interest in science

David Cyranoski, Tokyo

Four Nobel prizes and various initiatives to popularize science don't seem to be making much difference: a survey just released by Japan's science ministry says that the public is losing interest in science and technology.

Just 51% of the 2,084 adults polled earlier this year expressed an interest in science — a fall of 9% since the last time the survey was done in 1998.

Better results might have been expected in a country that nearly doubled its number of Nobel laureates in the sciences, from five to nine, officials at the ministry say. They are especially frustrated as they have tried to popularize science in various ways. ¥3 billion (US\$28 million) has been invested in the National Museum of Emerging Science and Innovation in Tokyo, which opened in July 2001, and science-based



television shows have been launched.

But according to the survey, which was released on 10 April, the number of people interested in what scientists have to say has been declining. In 1998, most of those who

said they would not attend public lectures on science said it was because they found them difficult to follow. In the latest poll, only one-third objected to the difficulty level — but almost 60% said either that they had no interest or that science did not seem relevant to their lives.

Noyuri Mima, a former cognitive scientist and the director of research and development at the new Tokyo museum, says she finds the results troubling. "Researchers get huge amounts of money but they do not try to explain what they do in a way people understand," she says.

Nonetheless, Mima is optimistic that innovative efforts such as the museum will pay off in the battle for the hearts and minds of the next generation. "We are trying to show the underlying process of science as it is happening," she says. "This will attract young people, but it will take time."

JAPANESE MINISTRY OF SCIENCE

Labelling laws for transgenic food come into effect

Laura Nelson, London

Europe this week introduced stringent rules for the labelling of food that contains genetically modified organisms. But in most countries the labels will take months to appear — and questions remain about how they will be implemented.

The rules, which are imposed by the European Union and came into effect on 19 April, are intended to aid consumer acceptance of genetically modified food in Europe. They may help to defuse a US complaint to the World Trade Organization (WTO) that Europe is unfairly blocking imports of transgenic food.

Food containing more than 0.9% genetically modified ingredients must be clearly labelled as doing so, the rules say. If the ingredients are awaiting final approval as being safe to eat, that threshold falls to 0.5%.

“This is the biggest piece of legislation in the food industry for 20 years,” says Richard Werran, head of Cert ID, a firm based in Fairfield, Iowa, that is offering to test food for its transgenic content.

Britain, Germany and the Netherlands are expected to implement the regulations in stores within a few months, but other nations may take longer.

The rules require food to be tracked from its source through manufacture to the point of sale. Manufacturers and packagers will also have to test food directly for traces of genetically modified organisms. A network of laboratories set up and operated by the European Commission (EC) has developed a series of standard tests.

The Institute for Health and Consumer Protection, part of the EC's Joint Research Centre in Ispra, Italy, has led the development of these tests, which use polymerase chain reaction technology to search for modified DNA.

But the tests do not work with the refined products, such as oil or sugar, of some genetically modified organisms because they may contain no transgenic DNA, so figures for food containing these will depend on manufacturers' supply-chain records.

Many retailers doubt that customers will buy food products labelled as genetically modified, and some refuse to stock them. Analysts are unsure whether the new rules will make any difference to consumer acceptance — or to the United States' complaint to the WTO. ■



Haul aboard: fertilizing the sea causes biomass changes that could be used to combat global warming.

Iron seeding creates fleeting carbon sink in Southern Ocean

Quirin Schiermeier, Munich

Dumping iron sulphate in the ocean to cause plankton blooms might not seem an eco-friendly way to tackle global warming. But, according to the most extended trial of the technique so far, it could prove an effective one.

The outcome of the trial in the Southern Ocean, which surrounds Antarctica, was published in last week's *Science*. It suggests that each atom of iron added to the sea could pull between 10,000 and 100,000 atoms of carbon out of the atmosphere by encouraging plankton growth, which captures carbon and sinks it deep towards the ocean floor (see *Science* 304, 417; 2004).

If successfully scaled up, such 'iron fertilization' of the sea could make a real dent in the high level of carbon dioxide in the atmosphere, which is causing global warming. Some researchers estimate that using the technique in the Southern Ocean alone could absorb 15% of carbon dioxide build-up. But ecologists caution that the technique could damage marine ecosystems in ways yet to be established (see *Nature* 421, 109–110; 2003).

A team of oceanographers from Californian marine research institutes dropped 1.7 tonnes of iron sulphate in the sea as part of the Southern Ocean Iron Experiment in 2002. They then used floating robots to measure the carbon flux — and found that lots of biomass was indeed created and consigned to the depths of the ocean, either as dead algae or fish excrement.

The findings have fuelled expectations that ocean fertilization could provide an environmentally friendly technique for reducing atmospheric carbon dioxide levels. “It is a worthy endeavour to mitigate future global warming,” says Russ George, chief scientist of the California-based Planktos

Foundation, a non-profit organization supported by the Canadian rock star Neil Young, which promotes large-scale iron fertilization.

Iron is essential for plant growth. Most of it reaches the oceans through winds carrying eroded, iron-rich soils from dry areas on land. But changes in climate and vegetation since the end of the last glacial period are believed to have diminished the iron supply to the ocean — reducing the growth of plankton, which naturally absorbs carbon dioxide from the atmosphere. Advocates of the technique argue that they would simply be restoring iron in the ocean to its previous level.

Researchers are, nonetheless, investigating the ecological risks of iron fertilization. The most recent trial, conducted in the Southern Ocean last winter, was the European Iron Fertilization Experiment. A team led by Victor Smetacek, a biological oceanographer at the Alfred Wegener Institute in Bremerhaven, Germany, is expected to report later this year on ecosystem changes resulting from the experiment's creation of an artificial plankton bloom.

The relation of carbon cycles to marine processes is complex and the authors of the *Science* paper say the jury is still out on the technique. “It would be premature to extrapolate our results to greater depths or larger areas,” says James Bishop, lead author of the paper and an oceanographer at the Lawrence Berkeley National Laboratory in California.

But Smetacek believes that human intervention on a grand scale in ocean chemistry is likely to happen, sooner or later. “For the first time we have the tools for large-scale operations with ocean ecosystems,” he says. “You will not do away with the mess that mankind has already made without using these tools. But we must face the challenge of using them responsibly.” ■

Lasers bend beams for desktop X-ray source

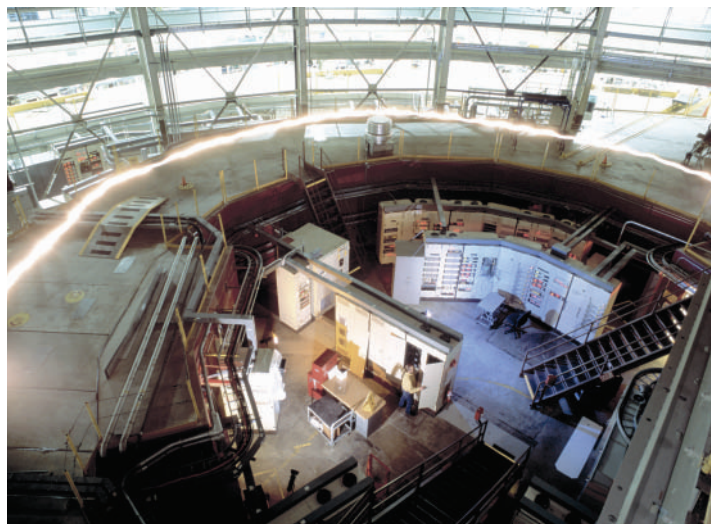
Jim Giles

The Advanced Photon Source in Argonne, Illinois, occupies an area the size of a sports stadium. But if Ronald Ruth is right, this tool, widely used in materials and biomedical science, can be shrunk down to fit on a desk.

Ruth's idea, which he outlined on 16 April, only exists on paper — for now. Yet it has already attracted \$7 million in support funding from the US National Institute of General Medical Sciences, which hopes that the device will help in the study of biologically important molecules. Ruth's background as an accelerator physicist at the Stanford Linear Accelerator Center in California also lends the project some credibility, synchrotron experts say.

Synchrotrons generate beams of high-energy X-rays, which can be used to determine the structure of proteins and materials. The beams are derived from electrons accelerated to within a whisker of the speed of light. These electrons are stored in large rings, some with diameters of hundreds of metres. Such devices are typically shared by hundreds of regular users, and demand for time on them often outstrips supply.

Ruth's alternative, launched at the Keystone Symposium on Structural Genomics in Snowbird, Utah, stores electrons at far lower energies — about 25 MeV compared with 7 GeV in the Argonne device — in a rounded rectangle just one metre long. The key difference is in the way it generates its X-ray beam.



Benchmark: laser manipulation of electrons could shrink the synchrotron.

X-rays are emitted when fast-moving electrons change direction. Originally, synchrotrons made use of the radiation generated as electrons were forced round the curved storage rings, but state-of-the-art devices use magnets to insert additional short deflections in the path of the particles.

Ruth intends to deflect the electrons using a fixed pattern of laser light, formed by bouncing a laser beam back and forth in a cavity that lies along the straight section of the mini-synchrotron's storage ring. This should bend the particles through tighter curves than is possible using magnets.

Ruth claims that this will produce X-rays with energies in the range of 5–35 keV, which should allow them to be used in many of the studies currently carried out at shared facilities. But the brightness of the beam will be orders of magnitude below that needed for

some experiments, such as work on the crystal boundaries below the surface of metals.

Some researchers say that they will believe Ruth's idea when they see it in action. Many designs look better on paper than they work in practice, says Janos Hajdu, a biochemist at Uppsala University in Sweden, who uses synchrotron radiation to study protein structure. "But I'll have one if he can make it happen."

The proposed machine, known as the Compact Light Source, will be marketed by Lyncean Technologies of Palo Alto, California, a company co-founded by Ruth. He is reluctant to put a price on the

device, but says it is likely to cost in the region of a few million dollars.

Although the notion of universities owning their own synchrotron is enticing, potential users say that the device could have drawbacks. Critics note that the slower-moving electrons will interact with each other and the walls of the device, limiting the time they can be stored, and for which X-rays can be produced, to a fraction of a second.

But Ruth says that electrons can be added to the ring while it is operating, allowing the device to produce a continuous beam of X-rays. If he is correct, Ruth's design might offer an advantage over other X-ray sources that use a laser to bend electrons, as, so far, these can only produce bursts of radiation. He adds that Lyncean is building a prototype that should start generating X-rays early in 2005.

Side effects leave smallpox vaccine in limbo

Erika Check, Washington

Trials of a new version of the smallpox vaccine have been halted because of a rare side effect, raising concerns about the vaccine's suitability for widespread use.

On 13 April, Acambis in Cambridge, UK, said that it had stopped recruiting patients into a large clinical trial of the vaccine because three people in the trial had developed myopericarditis — swelling of the heart muscle and surrounding tissue.

The US government has already ordered millions of doses of the new vaccine, called ACAM2000. But the latest finding makes it unlikely that it will be given to civilians unless there is an emergency.

The United States began a civilian

vaccination programme in 2002. So far, about 40,000 emergency workers have been given an older version of the vaccine to prepare them for a bioterrorist attack. But the programme has moved far more slowly than was intended, largely because of people's concerns about side effects.

Myopericarditis was not seen during the large-scale smallpox eradication campaign of the 1960s and 1970s, which used the version of the vaccine called Dryvax. But the side effect has emerged in new clinical trials of Dryvax. The US Department of Defense reports that, since December 2002, 77 of 615,149 military workers who took Dryvax developed myopericarditis. In the Acambis trials, which compared Dryvax with

ACAM2000, both versions of the vaccine seemed to cause the condition.

"It is still a big question whether this is something really new or whether these things happened before, but were not noted," says Anthony Fauci, director of the National Institute of Allergy and Infectious Diseases in Bethesda, Maryland.

Fauci says that ACAM2000 could still be given to civilians in an emergency. But health officials are unsure how to continue tests of the new vaccine. It is also not clear how efforts to license the vaccine will proceed. Fauci says that it may become necessary to stockpile a milder version of the smallpox vaccine — modified vaccinia Ankara — which is also being tested.



In the frame: palm civets on sale in China's markets may have been the source of SARS.

Foxes and shrews identified as potential carriers of SARS virus

Tokyo Foxes and hedge-shrews have been added to the list of animals that harbour the virus that causes severe acute respiratory syndrome (SARS) in humans. The list also includes cats and ferrets, although the masked palm civet remains the prime suspect as the reservoir for the virus.

China's official news agency said that some foxes and hedge-shrews in Guangdong, where SARS first emerged in November 2002, had tested positive for the coronavirus that causes the disease.

In early 2004, the virus was traced to wild-animal markets in Guangdong that sold palm civets, prompting a cull of up to 4,000 animals. But the new report casts some doubt on the link between civets and transmission to humans. The province's SARS research team found that market workers who handled civets were less likely to show clinical signs of being exposed to the SARS virus than other workers.

Gravity probe struggles to get off the ground

Washington A spacecraft designed to test Einstein's general theory of relativity missed its first launch window in California on 19 April. But several other windows, each only a second long, are planned for this week.

NASA's Gravity Probe B will orbit Earth's poles for about 16 months, looking for tiny gravitational distortions in the fabric of space, which Einstein's theory predicts should be caused by the mass of our planet. This test of general relativity will be 100 times more sensitive than any previous efforts, says Sasha Buchman, science mission manager for the experiment at Stanford University, California.

First proposed in 1959, the experiment has cost US\$700 million, and has been threatened with cancellation numerous times. To detect

the gravitational disturbances, the probe contains gyroscopes made with immense precision, which can measure movements of as a little as one hundred-thousandth of a degree. At the heart of each is a chunk of quartz the size of a ping-pong ball, milled to within 40 atomic layers of a perfect sphere. Each sphere is contained in a near-perfect vacuum with a pressure ten times less than the void of empty space. And the whole thing is cooled to -271°C to prevent the equipment being affected by heat. Initial results are expected in 2006.

California rejects plan for drug-producing rice

San Diego A proposal to begin commercial-scale growth of drug-producing transgenic rice in California suffered a setback last week when state officials declined to authorize the planting.

Ventria Bioscience of Sacramento had sought approval to plant 50 hectares of the rice on plots far from traditional rice fields in northern California (see *Nature* **428**, 591; 2004). The genetically modified rice produces the proteins lysozyme and lactoferrin, which can be used to treat a range of medical conditions. But officials at the state's food and agriculture department refused the request, saying that more public

input was required along with an additional federal permit.

Ventria is also seeking permission to continue growing smaller experimental plots of the rice this spring.

Database provides active link for human genes

Tokyo An international research team has pieced together one of the most comprehensive databases of human genes available. It contains information on more than 21,000 genes pooled from six existing collections of human complementary DNAs, the researchers report in a forthcoming issue of *PLoS Biology* (T. Imanishi *et al.* *PLoS Biol.* doi:10.1371/journal.pbio.0020162; 2004). That is two-thirds to three-quarters of the estimated total number of human genes.

Called the H-Invitational Database, the new resource carries information on active genes, including expression patterns in more than 60 human body tissues, predictions of the proteins they encode and the individual variations in genes that distinguish one person from another. Researchers can request many of the cDNAs by mail, and use them to study a gene's function themselves, says Takashi Gojobori, who led the project at the Center for Information Biology and DNA Data Bank of Japan in Mishima, Japan.

Engineers bang their barge for bridge safety

Washington Civil engineers this week are deliberately ramming a 635-tonne barge into a concrete support that formed part of an old bridge crossing Apalachicola Bay in Florida. But these are no carefree joy-riders — the stanchion is bristling with sensors that will gather data about the impact. Engineers believe that this will help them to design safer bridges.

The team plans to carry out about 12 collisions at speeds of up to 9 km per hour, and have already conducted seven crashes. "To the best of our knowledge, no one has done full-



scale tests of barges striking bridges," says Gary Consolazio of the University of Florida, who is lead engineer on the project. To see video footage of the crashes, visit www.nature.com/nsu/bridge

Although databases of genes are already available to scientists, this is the first time that active genes have been pulled together as one resource. The project took three years, and more genes will be added in the future.

♦ www.jbirc.aist.go.jp/hinv/index.jsp

Europe backs bid to tailor antidepressants

London The European Union has donated €9 million (US\$10.8 million) to a study seeking to improve antidepressant

prescription through genetic profiling.

The project, called GENDER, will follow about 1,000 depressed patients in 13 European centres as they take one of two common antidepressant drugs. The study aims to help clinicians diagnose and manage patients with depression.

Researchers, led by Peter McGuffin from London's Institute of Psychiatry, will study changes in gene and protein expression and look to see if certain molecules are associated with a particularly good or bad clinical outcome.



If your politics are green, you like your medicine 'holistic' and you're deeply worried by economic globalization, the chances are your fridge is full of organic produce. Today, support for organic farming is frequently part of a bigger social and political mindset — one that holds that 'natural' is best, and that naked capitalism is a threat to the health of the planet and its people.

But the origins of organic agriculture, in 1940s Britain, are more down-to-earth. Its pioneers were concerned, above all else, about the soil beneath their feet. Their philosophy was centred on practices designed to improve the richness and stability of the soil by restoring its organic matter and avoiding synthetic fertilizers, pesticides and herbicides. Wider concerns about biodiversity, social justice and animal welfare have grown from this core concept about how to manage our farmland's key resource.

These ideals have always set the organic movement squarely against intensive farming and chemical-based agribusiness. And, at least in public and in the media, those arguments rage more fiercely today than ever before. Yet behind the harsh rhetoric, a little-noticed convergence of views is taking place. For decades, the study of organic farming sat on the fringes of the green revolution in agriculture, as intensive techniques marched across the world, sending yields skyrocketing. But mainstream agronomists are becoming concerned about the long-term sustainability of this approach, and are focusing increasingly on soil integrity. Could it be that both sides of agriculture's great divide now want the same thing?

"It's been a huge move," says Mark Alley, an agronomist at Virginia Tech in Blacksburg.

ORGANIC:

Is it the future of farming?

In its pure form, maybe not. But elements of the organic philosophy are starting to be deployed in mainstream agriculture. *Nature's* reporters analyse this trend, assess the extent of organic farming worldwide, and frame the questions on which its wider adoption will depend.

"Twenty-five years ago, yield was everything. But in the past ten years, there's been a major recognition of the need to maintain organic materials in soil." And with the turn of the millennium, farmers have started to embrace approaches that keep soil structure intact and cut the high level of inputs — energy, fertilizer, pesticides and herbicides — that characterize intensive agriculture.

Going green ... again

These new methods diverge significantly from the purist organic vision. In particular, they rely heavily on 'low tillage' methods, which help to improve the soil but depend partly on the use of herbicides, fertilizers and pesticides. Those remain anathema to

the organic movement. But the change that is taking place — sometimes referred to as the second green, or doubly green, revolution — stems from a growing acceptance of the organic critique of the first one. Mainstream agronomists now acknowledge, for example, that intensive farming reduces biodiversity, encourages irreversible soil erosion and generates run-off that is awash with harmful chemicals — including nitrates from fertilizers that can devastate aquatic ecosystems.

For the organic movement, caring for the soil involves interspersing each harvest with a cover crop such as clover or rye that can fix nitrogen from the atmosphere. Cover crops keep down weeds, retain moisture and pre-



Organic principles are proving, at least in part, to be attractive to mainstream farmers, who have adopted a pesticide-free approach to controlling codling moth larvae (right).

vent erosion. Ploughing them into the soil at the end of the season restores the soil's organic content, and boosts its nitrogen content without the need to use synthetic fertilizer.

The low-till approach borrows heavily from these principles. Low-till farmers ensure that their soil is not left open to erosion by growing nitrogen fixers between rows of their cash crops and between seasons. But low-till farmers don't completely unhitch their wagon from conventional inputs. They still use nitrate fertilizers and pesticides as needed. Before each planting, they kill the previous crop with a broad-spectrum herbicide such as Roundup, made by Monsanto of St Louis, Missouri. This lets them punch the new seed directly into the ground through the decaying plants without tilling.

Till life

Low-till agriculture is taking root in both rich and poor countries. Pat Wall, head of conservation agriculture at CIMMYT, the International Maize and Wheat Improvement Center in Mexico, estimates that about 70 million hectares of arable land — something like 2% of the global total — is now using the method, with about a third of that in the United States.

Brazil has been in the vanguard of the change in the south, says Cheryl Palm, an ecologist specializing in tropical agriculture at Columbia University in New York. "It's swept through the country, and cut down soil erosion dramatically," she says.



Although many of the farms that are converting to low-till agriculture are large-scale operations, the approach is also rapidly gaining acceptance on smallholdings in places such as Ghana and India. On the Indian subcontinent, the area where low-till is being implemented has grown from nothing in 1997, through 100,000 hectares in 2001, to one million hectares this year.

Besides conserving soil structure, low tillage also reduces energy inputs. Farms in India that grow rice in the summer and wheat in the winter have cut their number of annual tilling operations from eight to one, Wall reports, reducing fuel use by 70%. "When I was there, the only people complaining about the change were the petrol station owners," he says.

Low-till farming also substantially reduces the need for chemical fertilizers. Cover crops provide some nitrogen initially, and then, as organic matter builds up in the soil, nitrates and other nutrients are less readily leached out of it, further decreasing

the requirement for added fertilizer.

But for organic purists, any approach to maintaining soil integrity that incorporates regular sprayings with Roundup and continued applications of nitrates is heresy. Peter Melchett, policy director of Britain's Soil Association, the world's oldest organic farming organization, scorns low-till approaches. "They tend to be something you can do for two or three years until you get grass weeds that aren't well-controlled by Roundup," he says. "Then you have to resort to ploughing."

Wall disagrees. He argues that the need for herbicide applications tails off after the first few years of low-tillage, as weed seeds disappear from the top layer of soil. "I think it is possible to get to no tillage and almost no herbicides," he says.

Such disputes might be resolved more readily if there was an abundance of data comparing pure organic methods with the low-till approach to soil conservation. "But there aren't a lot of long-term studies," says Mark David, a biogeochemist at the University of Illinois at Urbana-Champaign. "It isn't a simple comparison to make."

Chemical cuts

Some of the other ideas being borrowed from the organic movement — in particular a reduction in pesticide inputs — are resulting in a closer meeting of minds.

For instance, farmers have been forced to discard methyl bromide, the main soil fumigant that has been used to kill soil pests, as it will be phased out by 2005 under the Montreal Protocol to close the ozone hole. This has led farmers to experiment not only with other fumigants but with organic methods of killing insect larvae as well, including flooding fields between plantings and allowing the Sun to bake the soil through clear plastic sheeting.

Farmers are also bowing to consumer pressure. "People don't want pesticides in their food," says Diana Wall, director of the Natural Resource Ecology Laboratory at Colorado State University in Fort Collins. US orchards, for instance, have largely adopted organic methods for controlling the codling moth (*Cydia pomonella*), the larvae of which bore into ripening fruit and can destroy 80% of an apple crop without control.

In this case, concerns about pesticide residues on apples and pears led to legal restrictions on the use of organophosphates, the most effective class of pesticides, under the 1996 Food Quality Protection Act. Organic control of the codling moth disrupts mating by releasing sterile males and spraying female sex pheromones to confuse the rest.

The doubly green revolution doesn't necessarily embrace the broader aspects of organic ideology, such as social justice and animal welfare. But if the organic movement wants to change the world, it is making a reasonable start.

Colin Macilwain

E. THOMPSON/AP

RUNK/SCHOENBERGER/GRANT HELLMAN PHOTOGRAPHY



ORGANIC FAQs

In the developed world, sales of organic produce are growing rapidly. But how far can this trend extend? That depends on how strictly you define organic farming ... and the answers to three other pivotal questions.

What is organic farming?

At the core of the organic philosophy lies a ban on the use of synthetic fertilizers, pesticides and herbicides. That means adopting other techniques to nourish crops and protect the soil, such as growing 'cover' crops between seasons to prevent erosion and to restore organic matter.

The organic movement also encompasses such tenets as animal welfare, energy efficiency, social justice and the simple agrarian ideal of small farms growing produce for local communities. It is on this last point that the success of organic farming is starting to divert the movement from its pure vision.

Although organic produce remains a niche market, global sales have risen by about 20% per year for five years running¹. This growth has seen some 'organic' producers turn into industrial-scale ventures that ship their products over thousands of kilometres. Organic proponents may aim to reduce the fossil fuel expended in transporting crops by encouraging farmers to sell to local markets,

but the popularity of organic food in wealthy countries has spawned a huge export market. North America and Europe account for 97% of global organic food and drink sales, but nearly half of the world's organic farmland is found in Asia, Australia and Latin America (see Map, page 794). It's hardly what the movement's pioneers had in mind.

Organic standards also differ in their details from country to country. Most rules governing organic farming, including those laid down in the European Union, Japan and the United States, are based on standards set by the non-profit International Federation of Organic Agriculture Movements in Bonn, Germany. "Anyone who is really credible will adhere," says Bruce Pierce, deputy research director of the Elm Farm Research Centre in Berkshire, UK, which studies methods of organic cultivation. But standards are not always comprehensive: Japan, for instance, has no rules for organic meat.

And there are important differences between regions of the world. For instance,

an American farmer who chooses to use Chilean nitrate, a mined source of sodium nitrate, permitted under US rules, could not sell the resulting produce in Europe. Although Chilean nitrate is a natural substance, European organic standards consider it to be the equivalent of a synthetic fertilizer because it is highly soluble and leaches readily from the soil. Nor could the US farmer market milk from a cow that had been raised on an organic diet for less than a year — European rules are stricter than US standards, which require only six months.

Similarly, a consumer buying organic produce in the United States cannot be sure that it is free of contamination from genetically modified (GM) crops unless this is explicitly stated on the label. In Europe, GM content in all GM-free food, including organic produce, is limited to 0.9%, and some certifying bodies, such as Britain's Soil Association, allow no detectable GM.

Because organic products fetch premium prices, there are concerns about the possibility of cheating. Organic rules are enforced by farm inspections, but the logistics can be difficult, particularly in remoter parts of exporting countries. "The inspection process is not completely foolproof," admits Francis Blake, standards director of the Soil Association and a former inspector. "It relies on trust."

As a result, some researchers have begun to look for ways to test organic products for authenticity. Alison Bateman of the University of East Anglia in Norwich, UK, is developing a test based on the higher proportion of the isotope nitrogen-15 in organically farmed soil. This is because nitrogen-fixing plants accumulate more of this heavy isotope than is present in synthetic fertilizers. Other researchers are investigating whether concentrations of trace elements such as calcium, boron, magnesium and selenium differ between organic and conventional produce.

But tests such as these address only the final products, and so cannot verify whether the farm from which the food came adhered to the principles of organic agriculture. "They cannot tell you if a product has been organically managed or not," says Blake. **Laura Nelson**

Is organic food better for us?

This is the claim that attracts many of the consumers who buy organic, so it's no surprise that the movement's advocates answer with an unequivocal 'yes'. In 2001, for instance, the Soil Association concluded unambiguously that organic food contains less of the bad stuff, such as pesticides, and more of the good stuff, such as vitamins and minerals².

But independent scientists are less convinced. They say that many comparisons between organic and conventional produce are let down by poor methodology. For

example, some studies fail to take into account the fact that organic farmers prefer crop varieties that are resistant to disease, whereas conventional farmers focus on high-yielding strains. Such studies confuse the effect of production system with variety.

Apparent benefits may also turn out to be superficial. Several studies have shown that organic crops contain higher levels of nutrients such as vitamin C and iron³, for example, but most people in developed countries already have enough of these compounds in their diet. "When evaluating relative nutritional value, these are not important targets," says Kirsten Brandt, an agricultural scientist at the University of Newcastle upon Tyne, UK.

On the other hand, plant secondary metabolites, substances that may be present at higher levels in organic food, could be an appropriate target, says Brandt. Phenolic metabolites, which fruit and vegetables produce to ward off insects, are believed to have anticancer properties, for example. Last year, food scientist Danny Asami and his colleagues at the University of California, Davis, looked at organic and conventionally grown marionberries and maize (corn) from

the same farm, and found 30–50% more phenolics in the organic samples⁴. Studies of organic pears and peaches have also showed raised levels of phenolics⁵.

Brandt, who tracks such studies, says that the evidence points to organic crops containing 10–50% more secondary metabolites than conventional equivalents. This may be because fertilizers applied to conventional plants supply a surfeit of nutrients, encouraging the plant to channel more energy into growth, rather than defending against pests.

But do plant secondary metabolites really do us any good? Anthony Trewavas, a plant scientist at the University of Edinburgh, UK, and a high-profile critic of the organic movement⁶, questions whether we should be trying to boost levels of secondary metabolites before we know the answer to this question. About 10,000 of these metabolites are thought to exist. Many that have been studied seem to behave paradoxically, acting as carcinogens at high doses and showing anticancer properties at low doses⁷.

At the very least, it seems reasonable to expect organic food to be free from pesticides, which are banned or severely restricted under organic regimes. Food-safety authorities monitor pesticide residues in conventional crops, but levels do occasionally exceed maximum legal limits. And according to a 1998 study by Britain's Consumers' Association, some pesticides remain on fruit and vegetables even after they have been washed⁸.

But should we be worried about this? Most researchers believe that allowed residues are safe, although uncertainties exist. The difficulty in applying results from animal experiments, which are the mainstay of toxicological assessments, to humans is one

problem. Opinions can also be revised. The chlorine-containing pesticide lindane was banned in Europe in 2001, for example, because of concern that it might promote breast cancer⁹. Ultimately, most toxicologists urge caution in assessing and regulating pesticide residues, but they don't see the need to eliminate them entirely.

Jim Gilles

Is organic farming better for the environment?

This is a more complex question than it at first appears. In some arenas, such as biodiversity, organic farming has clear benefits. But in others, such as runoff and atmospheric emissions, the differences between the two systems are difficult to establish.

Although few large, long-term studies directly comparing the systems exist, several literature surveys have brought together smaller studies to build overall comparisons^{10,11}. There is general agreement on some benefits. For example, organic farms do better than conventional farms at nurturing abundant and diverse populations of plants, insects and other animals. And organic farms release no synthetic pesticides or herbicides, some of which have the potential to harm wildlife.

Organic farms also score points for using less energy — both per unit area and per unit of yield — and producing less extraneous waste, such as packaging materials for chemicals and feed. A typical study at Washington State University in Pullman totted up the energy consumed by labour, machinery, electricity, fertilizer, pesticides and weed control to grow apples in organic and conventional orchards, and found the organic orchard to be 7% more energy efficient¹².

On the flipside, organic methods have a greater environmental impact in some small ways, studies show. Methane emissions from organic farms are likely to be higher per unit of food production, for example. At least in the United States, where dairy cows receive growth hormone, organically raised cattle yield considerably less milk than their hormone-assisted peers — requiring more cows, which collectively pass more methane.

But findings are less definitive about the much more significant environmental impact of farm runoff — through which nitrates and phosphorus leach into streams, rivers and lakes, causing algal blooms that suffocate fish. Several studies have suggested that organic methods will reduce nitrate leaching, but according to a 2003 assessment of the literature sponsored by the British government, the various factors that affect runoff mean that this is not guaranteed¹³. Too few measurements of phosphorus runoff have been made to determine which system releases less, the report concluded.

In theory, organic farms are friendlier to the atmosphere. They should, for instance, generate less carbon dioxide, which is released

P. DEAN/GARNT HEILMAN PHOTOGRAPHY



Where there's life: the broad biodiversity supported by this organic cereal crop is clear to see.

F. GILSON/STILL PICTURES

in abundance in conventional farming by burning fossil fuels to manufacture, transport and spread nitrate fertilizers. And the ploughing into the soil of crop residues and cover crops should pull carbon back out of the atmosphere more efficiently. Organic methods might also be expected to produce less nitrous oxide — one of the causes of acid rain — than is released by heavily fertilized soils.

Although the British assessment found that organic farming does lead to lower CO₂ emissions, it also said that a lack of firm data made it impossible to compare emissions of nitrous oxide — which is also produced by legumes and manure on organic farms. Nor were there enough data to evaluate the effectiveness of the two systems as sinks to capture atmospheric carbon.

For organic advocates, the key environmental issue is not the year-by-year balance of farming inputs and outputs, but rather the long-term sustainability of the system. By recycling both nitrogen and organic material back into the soil, they believe, organic agriculture can ensure this.

Many studies support the idea that organic methods are good for soil quality¹⁴. “I used to be sceptical about organic methods, but the evidence on organic material changes things,” says Mark David, a biogeochemist at the University of Illinois at Urbana-Champaign. But in the absence of long-term comparative studies, the argument about sustainability is difficult to prove.

Colin MacIwain

Can organic farming replace conventional agriculture?

Not if the world wants a meat-rich diet, as even die-hard organic proponents are willing to concede. But if the world's demand for cheap, abundant meat can be curbed, then quite possibly it could.

Ultimately, it's a question of efficiency and yield. The elimination of pesticides and herbicides does not seem to reduce yields as much as you might expect. Because pests tend to prefer particular plants, the crop rotations favoured by organic farmers help to prevent insect populations from accumulating to damaging levels. Continuous cover cropping in winter also keeps weeds down, so the soil accumulates fewer weed seeds. Natural pesticides and mechanical weeding finish the job.

Still, there are some regional pest problems for which no organic solution has yet been found. Notorious in the northwestern United States is the garden symphylan (*Scutigerebella immaculata*), a centipede that can attack asparagus, maize, mint and strawberries and is controlled by soil fumigants in conventional systems. Years of research have yielded no effective organic control — all organic farmers can do is replough the field in an attempt to kill the centipedes.

A bigger influence on yield is the means by which organic fields are supplied with enough



Soil survivor: the garden symphylan has so far proved resistant to any organic pest controls.

nitrogen to maintain productivity. Conventional fields get a generous dose of nitrogen each season from synthetic fertilizer, whereas organic fields get theirs from manure and cover crops, sometimes called ‘green manure’.

Season to season, the two approaches can produce comparable yields. A 21-year study by the Research Institute of Organic Agriculture in Frick, Switzerland, concluded that organic fields produce yields 20% lower than conventional fields, on average¹⁵. Meanwhile, another long-term study by the Rodale Institute in Kutztown, Pennsylvania, obtained roughly equal yields of maize and soya beans with the two systems. The Rodale team also found that organic systems can achieve 20–40% higher yields in drought years¹⁶.

But to maintain the soil's nitrogen content in the long term, organic farmers must grow a legume or other nitrogen-fixing crop regularly. This takes land out of commercial production, reducing the overall yield of a plot over time. Although some legumes are edible, the most efficient nitrogen-fixers, such as clover, are not. The Rodale researchers managed to minimize lost yield by growing legumes over the winter, but this may not be practical in harsher climates; nor will it provide the same benefit in warmer regions where cash crops are grown year round.

Vaclav Smil, a natural-resources researcher at the University of Manitoba in Winnipeg, Canada, calculates that there is an even bigger obstacle to organic edging out conventional farming any time soon. He says that the total nitrogen available to organic farmers through manure and legumes amounts to less than half the total nitrogen consumed by the world's farms today — some 85 million tonnes. More cover cropping may increase the available nitrogen, but this is a luxury farmers in places such as Indonesia or China can ill afford. “In these countries, you cannot plant crops that no one will eat,” Smil says.

In large part, the huge nitrogen inputs required by modern agriculture are needed to grow sufficient grain to raise livestock: producing a kilogram of lean meat requires 25–50 kg of grain. Even with sufficient nitrogen, organic farms would have a hard time meeting this demand, Smil says, because they must grow a variety of crops to maintain soil health and defend against pests. Turning over entire farms to grow maize and soya beans to feed livestock isn't a viable option.

Even stalwart supporters of organic agriculture agree. But their line is that we should eat less meat and more vegetables, and embrace an organic future. The argument that organic farming can't produce enough food “only works if you assume that we continue to expand production of cheap meat”, says Peter Melchett, policy director of Britain's Soil Association.

Virginia Gewin

1. Willer, H. & Yussefi, M. (eds) *The World of Organic Agriculture: Statistics and Emerging Trends* (International Federation of Organic Agriculture Movements, Bonn, 2004).
2. Heaton, S. *Organic Farming, Food Quality and Human Health: A Review of the Evidence* (Soil Association, Bristol, 2001).
3. Worthington, V. J. *Altern. Complement. Med.* **7**, 161–173 (2001).
4. Asami, D. K., Hong, Y.-J., Barrett, D. M. & Mitchell, A. E. *J. Agric. Food Chem.* **51**, 1237–1241 (2003).
5. Carbonaro, M., Mattera, M., Nicoli, S., Bergamo, P. & Cappelloni, M. *J. Agric. Food Chem.* **50**, 5458–5462 (2002).
6. Trewas, A. *Nature* **410**, 409–410 (2001).
7. Ames, B. N. & Gold, L. S. *Mutat. Res.* **447**, 3–13 (2000).
8. *Health Watch* 8–11 (June 1998).
9. Mitra, A. K., Faruque, F. S. & Avis, A. L. *J. Environ. Health* **66**, 24–32 (2004).
10. Stolze, M., Piorr, A., Häring, A. M. & Dabbert, S. *Environmental Impacts of Organic Farming in Europe* (University of Hohenheim, Stuttgart-Hohenheim, 2000).
11. Hansen, B., Alroe, H. J. & Kristensen, E. S. *Agric. Ecosyst. Environ.* **83**, 11–26 (2001).
12. Reganold, J. P., Glover, J. D., Andrews, P. K. & Hinman, H. R. *Nature* **410**, 926–930 (2001).
13. Shepherd, M. et al. *An Assessment of the Environmental Impacts of Organic Farming* DEFRA-funded project OF0405 (DEFRA, London, 2003).
14. Johnston, A. E. *Soil Use Manage.* **2**, 97–105 (1986).
15. Maeder, P. et al. *Science* **296**, 1694–1697 (2002).
16. Lotter, D. W., Seidel, R. & Liebhardt, W. *Am. J. Altern. Agric.* **18**, 146–154 (2003).

Statistics don't support cot-death murder theory

Misunderstanding of statistics is widespread and has led to miscarriages of justice.

Sir—Your News story “Quashed convictions reignite row over British cot deaths” (*Nature* 427, 384; 2004) notes concern over the testimony of an expert witness, the distinguished paediatrician Roy Meadow, in several UK court cases in which parents were convicted of murdering their children.

Another aspect of his testimony that should be discussed more widely is the use of statistical arguments to conclude that the probability of two children in the same family dying a cot death was 1 in 73 million. Meadow based this testimony on statistics given in a government-commissioned

report. The ‘1 in 73 million’ statistic was not the sole basis for Meadow’s expert testimony, but sadly, neither the defence nor anybody else in court challenged the simple assumption behind the calculation that two cot deaths in the same family are unlikely to be related in any way.

A more convincing analysis by Ray Hill, a mathematician at Salford University, UK, reveals that, if there has already been one cot death in a family, the chance of a second one is 10 to 22 times higher. See <http://pass.maths.org.uk/issue21/features/clark>, where Helen Joyce uses Bayes’ theorem to calculate the probability that

the deaths arose from natural causes, by using plausible values for the alternative hypothesis that they were murdered. This analysis, which does not require an understanding of the underlying causes of cot death or child murder, leads to a probability of 0.625 that the deaths were natural (see B. Lewis, *Math. Gaz.* 87, 418–431; 2003).

Unhappily, the understanding that statistics is a difficult subject is not widespread, even among distinguished paediatricians.

Hermann Bondi

Churchill College, Cambridge CB3 0DS, UK

Dangers of crying wolf over risk of extinctions

Sir—Media coverage of conservation research is usually welcomed by the scientists involved, but there are pitfalls to heed. Damaging simplifications of research findings may expose conservationists to accusations of crying wolf, and play directly into the hands of anti-environmentalists. For example, in January 2004 it was widely reported in the UK print media that one million species would go extinct by 2050. The original report (*Nature* 427, 145–148; 2004), however, was based on 1,103 species and clearly stated that — as a consequence of climate change over the next 50 years — a variable proportion of land animals and plants might eventually go extinct.

We reviewed 29 reports in the local and national UK press, and found that many of the errors could be traced back to the press releases and agency newswires. In a press release from the lead author’s university, the figure of a million species appears along with the claim that a quarter of all land animals and plants may go extinct — but *eventually*, not by 2050. Newswires ranged from the cautious (“Hundreds of species of land plants and animals around the globe could vanish or be on the road to extinction over the next 50 years if global warming continues” — Dow Jones International) to the sensational (“Global warming could wipe out a quarter of all species of plants and animals on earth by 2050” — Reuters).

Unsurprisingly, subsequent newspaper articles in the national and local press were highly inaccurate: 21 of the 29 reports we reviewed claimed that a million or more species would be extinct by 2050. Two reports even claimed that one-third of the entire world’s species would become

extinct. No reports specified the full range of uncertainty (5.6% to 78.6% of the species studied would be committed to future extinction) and only two correctly stated that most species would become extinct well after 2050 (full details of our survey can be seen at www.geog.ox.ac.uk/research/biodiversity/pubs/index.html).

Politicians and conservationists repeated these statements. The European Union’s environment commissioner Margot Wallström, for example, commented on “the recently published study that suggests global warming could wipe out a third of the planet’s species by 2050”.

How can the conservation community prevent a repeat of such wide-scale media misrepresentation? Practical steps might be for high-profile journals to restrict press releases in the climate-change arena to research papers that present clear and unequivocal findings, and for scientists to write to newspaper editors and politicians to clarify misleading media articles. More generally, any institute, journal or individual involved in putting out a press release has a responsibility to ensure that it is both accurate and perfectly clear.

Richard J. Ladle, Paul Jepson, Miguel B.

Araújo, Robert J. Whittaker

Biodiversity Research Group, School of Geography & the Environment, Oxford University, Mansfield Road, Oxford OX1 3PS, UK

Error message

Sir—In the 19 February 2004 issue of *Nature*, there were ten items (one Brief Communication, one Article and eight Letters to *Nature*) containing figures with error bars, but only three had figure legends describing what the error bars were: in one case, 80% confidence intervals; in another, standard deviations;

and in the third, standard error of the mean. The articles with incomplete legends represented both the biological and physical sciences, across many different disciplines, and clearly should not be considered isolated examples.

Error bars can be used by the reader to determine how much the data varied, allowing an estimation of whether the experiments gave reproducible results, whether the results were significantly different from the controls, and sometimes whether the data were obtained in an unbiased manner.

How did these omissions occur? If authors include error bars on their figures, why do they so often forget to state what they are in the legends? How can reviewers be confident that the conclusions are correct if they are not told about the errors in the data? Why don’t reviewers request that descriptions of the error bars be included when they review the papers?

When properly described, error bars can be very revealing. In their analysis of the experiments and methods used by Jacques Benveniste to study homeopathy, John Maddox and colleagues illustrated how much information can be gained if one knows how to interpret errors correctly (*Nature* 334, 287–290; 1988).

By not ensuring that all papers that have error bars describe what they are, *Nature* publishes material that cannot be properly assessed by its readers.

David L. Vaux

The Walter and Eliza Hall Institute, 1G Royal Parade, Parkville, Victoria 3050, Australia

***Nature* is fortunate in having such attentive readers. Our editors and reviewers expect error bars to be properly defined, and we shall be more vigilant in ensuring best practice in future — Editor, *Nature*.**

The miracle of the mould

Howard Florey and colleagues overcame great obstacles to isolate penicillin.

The Mould in Dr Florey's Coat: The Remarkable True Story of the Penicillin Miracle

by Eric Lax

Little, Brown: 2004. 288 pp. £16.99

William Shaw

Well-researched and readable accounts of medical science and disease are always welcome. The first thing to note about this new work by Eric Lax, whose earlier efforts have been a well-regarded biography of Woody Allen and an engaging account of cancer chemotherapy, is that its title promises a great deal — and does so rather late in the day.

There is little doubt that, after more than half a century of personal reflections and scholarship, the story of the emergence of penicillin as a life-saving medicine remains remarkable — not least for the obstacles overcome and for the personalities of the trio of Nobel laureates involved: Alexander Fleming, Howard Florey and Ernst Chain. But promise of a true story begs the question of what has been on offer since 1945. And was there a miracle? Well, nearly, at least in the sense that the small team assembled by Florey at the Sir William Dunn School of Pathology in Oxford, UK, in the dark and difficult early years of the Second World War, hardly seemed like a research unit likely to isolate and purify a substance active at very high dilution and already known to be unstable. There were also long odds on their demonstrating its safety and efficacy, as well on their ability to marshal pharmaceutical forces in a drowsy America — while German bombs were falling on Britain each night.

Two items stand out in Lax's introduction, the first being a brief account of the case of Anne Miller, treated successfully in March 1942 at New Haven Hospital for streptococcal sepsis with penicillin from Merck, the first systemic cure in the United States and a trigger to accelerated production there. It was her obituary in the *New York Times* 57 years later that sparked Lax's interest. The second item, which flags a story that he develops, reads: "There are four people at the heart of the penicillin story: Fleming, Florey, Heatley, and Ernst Chain."

Norman Heatley's critical role in the proving of penicillin was not recognized until 1990, when Oxford awarded him an



Norman Heatley (below left) oversaw the mass production of penicillin in 1940s America.



honorary doctorate of medicine, the first non-medical person to be so honoured. Although several authors have sought Heatley's recollections, few have had access to his notes and diaries, which cover the early years of the penicillin story. Lax is the first to put Heatley's perceptions front and centre, from long interviews with him and careful scrutiny of his writings. That Lax felt com-

pelled to do so owes less to a deliberate lack of emphasis on Heatley's role by earlier writers than it does to Heatley's self-effacement. In fact, the two definitive biographies of Florey, by Gwyn Macfarlane and Trevor Williams, give Heatley much credit, although Ronald Clark in his *Life of Ernst Chain* focuses more on his larger-than-life title character.

In addition to drawing heavily on a dozen or more standard references on the early history of penicillin, including those noted above, Lax has made good use of a wide range of primary material. This includes UK sources, such as the Florey archives at the Royal Society, Chain's papers in the Wellcome Library and those of Edward Abraham in Oxford's Bodleian Library, and archives

from farther afield, such as the Yale papers of John Fulton, Florey's close friend and champion, the archives of the Rockefeller Foundation in New York, and the Nobel archives in Stockholm. The author's efforts to explain biological or chemical concepts or procedures to non-scientists are less successful, however, and occasionally wrong. But this detracts little from an engrossing story and is unlikely to lead lay readers astray.

Although *The Mould in Dr Florey's Coat* is hardly a fresh look at a familiar subject, it does break new ground in providing a more balanced view, one that attempts to reconcile competing claims, either of the principals or their biographers. There are splendid opening chapters on each of the four players: "The Quiet Scot" (Fleming), "The Rough Colonial Genius" (Florey), "The Temperamental Continental" (Chain) and "The Micro Master" (Heatley). Then the narrative begins, slowly at first, accelerating as penicillin begins to arrive from America — a trickle initially, but in sufficient amounts by 1943 to reach military surgeons in the field, with Florey in North Africa advising on its use.

But the best of the story covers the early years from 1939 to late 1942, when nothing was certain, including the success of Florey and Heatley's journey to the United States to seek support for the industrial production of

penicillin. Perhaps it was then that *Penicillium notatum* was in Florey's coat pocket, or even, from Heatley's account of a practice he suggested earlier, present as spores rubbed on the lapels of their jackets for safekeeping.

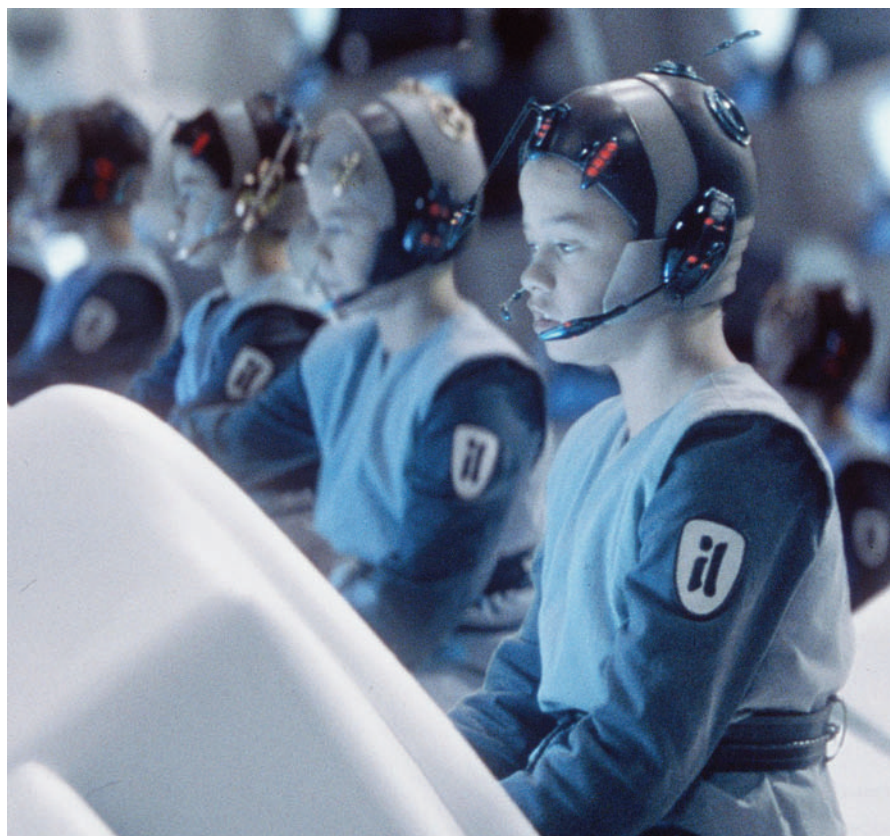
The virtue of Lax's approach is that it focuses not only on members of the Oxford team and their relationships with one another but also on the great and the good of Britain and America, who in one way or another helped to make it happen, notwithstanding frequent lapses in their comprehension of the staggering difficulties facing Florey. Lax helps modern readers understand that the 'proof of principle' mouse experiments took place as Dunkirk was being evacuated, and that the consequent push to enhance the supply of penicillin in Oxford coincided with the onset of the Blitz.

Florey's complicated personal life from 1940 onwards is also part of the story. With their children evacuated to the United States from 1941, Ethel and Howard Florey struggled to work out how to remain together while Howard continued his demanding professional life, Ethel found ways to live with deafness and contribute to clinical work with penicillin in Oxford, and — not part of the agreement — Howard turned to his laboratory associate Margaret Jennings for affection, said by one observer to be "one of the worst-kept secrets in Oxford".

There is ample evidence to support Lax's decision to bring Heatley to the foreground, closer to Florey than to Chain. Florey not only valued Heatley's unquestioned experimental skills, gift for improvisation and intelligence, he trusted and depended upon him. Heatley's respect for, and commitment to, "the Professor" were unqualified. Lax makes it clear that the chemistry of their relationship enabled Florey to focus on the big issues, channelling his energy and broad intelligence into important pursuits. Thus, when he could leave nothing to chance, Florey turned to Heatley, rather than Chain, for his travelling companion on the Pan American clipper to New York from Lisbon in July 1941, to join him for virtually every meeting in America with industry or government figures, and leaving him behind for a full year to assist and advise the parties who would undertake the large-scale production of penicillin.

Sadly, Heatley never had the opportunity to read Lax's account, as he died on 5 January 2004, just before his 93rd birthday. Had he still been alive, one can imagine him objecting that far too much attention had come his way of late, that luck had always played a large role in the enterprise, and that he agreed with Pasteur: being prepared to recognize good fortune, and having the wits to act upon it, was sufficient. ■

William Shaw is professor emeritus of biochemistry and chemical microbiology at the University of Leicester, UK. e-mail: bill.shaw@rcn.com



A real risk? Cloning is used to create warriors in *Star Wars Episode II: Attack of the Clones*.

A clear view of cloning

A Clone of Your Own? The Science and Ethics of Cloning

by Arlene Judith Klotzko
Oxford University Press: 2004. 162 pp.
£12.99

John Harris and Tuija Takala

For members of the general public confused by the heated debate generated by human cloning, *A Clone of Your Own?* does an admirable job in explaining many of the complicated ethical and scientific issues without giving way to academic jargon. Drawing from literature, the visual arts, films and her personal experiences, Arlene Judith Klotzko has created a comprehensible overview of cloning.

Klotzko tells us about the early experiments that Aristotle did with chicken embryos; about a German scientist, Hans Spemann, who figured out the theory of cloning some sixty years before we actually succeeded in cloning mammals; about Dolly, the first cloned sheep, and other work done at Edinburgh's Roslin Institute; and the current successes and failures in attempts to clone mammals. She explains the basics of embryology and provides an overview of stem-cell research.

The book paints a picture of human

cloning as a worthwhile enterprise. Klotzko lists the various potential benefits that therapeutic cloning could have, and explains why we might have reservations about reproductive cloning, even if it should not be banned completely. She also, following several others, clarifies many of the common misconceptions concerning the identity of clones, and puts in plain words the restricted effect that genes have in shaping the kind of people we become.

For a book with radical pretensions, Klotzko's arguments are restrained in their defence of human reproductive cloning — she is impressed by the supposed risks entailed. One of the major objections to any current attempt to clone a human is that, in the case of Dolly, only one clone was successfully produced after 277 attempts. Cloning is inefficient and wastes many embryos. But embryo wastage cannot be an objection to reproductive cloning for those who accept natural reproduction — after all, about 80% of embryos perish in natural reproduction. For every live birth, 3–5 embryos are created only to die.

We must remember that giving birth in the normal way isn't safe for the mother or the child. It is so risky that early abortion is safer for the mother than childbirth. Moreover, 3–5% of babies born have some abnormality. So the safety of reproductive cloning is at best a contingent argument that fails utterly if cloning could be made safe. Furthermore, the safety argument is interestingly

Science in culture

Platonic puppetry

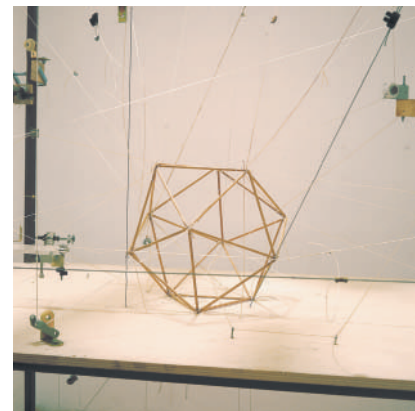
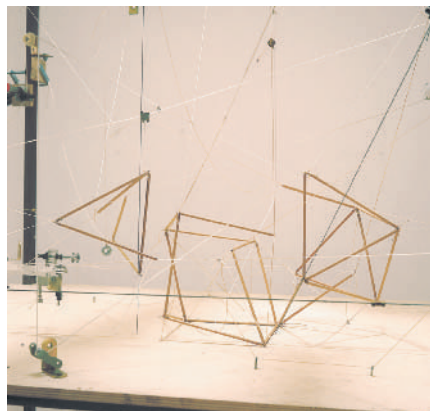
Attila Csörgő's kinetic sculptures bring regular polyhedra to life.

Martin Kemp

"Bringing order out of chaos" is a common enough phrase. It is applicable to any act of conspicuous resolution of what was previously a muddle. But in science and art it has more profound resonances. Many scientific experiments and many works of art rely on the setting up of conditions for discerning underlying order, first apparent to the scientist or artist, and then to ourselves as surrogate witnesses. A key difference, historically, is that science has repeatedly used time to set ordering processes in motion, either through the plotting of process or the dynamic achievement of experimental ends. The visual arts, by contrast, have been restricted largely to static evocations of order and disorder.

With the advent during the twentieth century of art that actually moves — sculptures in constant motion or images that exploit the technologies of video and computer graphics — this traditional limitation no longer applies. But it is rare to find motion in art used cyclically to choreograph, first, the transformation of orderly structures into intermediate states in which an inherent order is no longer discernible, and second, the reconstitution of a new regularity which is different from that of the original. Such transformational processes lie at the heart of the spectacularly intricate sculptures by Hungarian artist Attila Csörgő.

The dimension of time is embedded in all his work, whether it involves the rotational imprinting of an all-round image on a hemisphere by a spiralling camera, or devices in which apparently inchoate elements assume legibility under rotational motion. His most famous works, which first attracted international renown at the Hungarian pavilion in the Venice Biennale in 1999, exploit extraordinary mechanical contraptions to dismember and reconstruct the platonic solids — the five regular polyhedra that Plato believed to be



Taking shape: Attila Csörgő's geometrical figures are constructed and taken apart in real time.

the shapes of the fundamental components of the physical universe.

In the middle of a racked apparatus of electric motors, strings, pulleys and weights sit the skeletal shapes of a tetrahedron, an octahedron and a cube, composed, respectively, of 6, 12 and 12 wooden batons. As the apparatus whirs into motion, the cat's cradle of strings pulls the rods apart, deconstructing the geometrical figures. The rods wheel into space like the disarticulated limbs of broken puppets. If we could trace the tracks of the ends of the rods, we could mentally wind back time to realize the order still distantly immanent in the array. What happens next is not an exact reconstitution, however. The components twist unerringly to settle, end on end, into one of the more complex solids — a dodecahedron in one of the apparatuses, and an icosahedron in another. Momentarily resolved, the mechanisms then embark on the reverse direction of the cycle of dissolution and crystallization.

The process is beguiling, like a musical composition reaching a resolution that seems inevitable yet remains surprising. The visual quality even has

a sensual dimension — a quality suggested by the title of Csörgő's new exhibition, 'Platonic Love', which is at Kettle's Yard, University of Cambridge, until 9 May. He is tapping into the enduring aesthetic of the geometrical bodies, which has long fired cosmologists of a keplerian bent.

The geometrical results could, of course, have been achieved through computer graphics, but the visibility and evident physicality of the mechanisms are integral to the spectator's engagement. Our fascination with the raw mechanics of the process is akin to our continued delight in simple optical illusions, even in an age overloaded with film, TV, video and computer animations.

Csörgő, like a scientist, presents us with an anatomized vision of what lies inside dynamic phenomena, of how an array of no discernible order can be coherently characterized within a temporal frame as a moment in a constant flux between two oscillating states of resolution.

Martin Kemp is professor of the history of art at the University of Oxford, Oxford OX1 1PT, UK, and co-director of Wallace Kemp Artakt.

♦ www.kettleyard.co.uk/exhibitions/csorgo

two-edged. If any reproductive technology, including cloning, could be made safer than normal sexual reproduction (as may well eventually happen), then those who regard safety as decisive would have to abjure sexual reproduction for its safer technology-based counterpart.

The great promise of cloning in terms of human welfare, however, lies in the use of these techniques not for reproduction but for therapeutic purposes. The regenerative properties of stem cells that make them so attractive as a possible therapeutic tool also mean that the distinction between therapy and enhancement will inevitably be further eroded. Treatments that cause tissue to repair itself *in situ* and go on doing so are likely to extend lifespan. If therapies are

developed that modify cells to be resistant to cancer or HIV/AIDS, this will create unprecedentedly enhanced humans. Anyone who is disturbed by such a prospect will have the most agonizing of choices to make if the promise of stem-cell research is fulfilled.

A Clone of Your Own? provides engaging and clear explanations of both the basic scientific issues and related ethical issues surrounding cloning. Klotzko appears to have drawn on a wide range of published work on the ethics of cloning, and makes a large proportion of the arguments in the literature accessible in this short book. However, readers who are unaware of the literature may be left with the impression that Klotzko is the first, and almost the only,

person to have written on the ethics of cloning, which is far from the case. It is a pity that the author and the publisher have provided so little reference to the extensive ethics literature and given so little sign that they are even aware of it. Whether one's interest lies in the science or the ethics of cloning, the short list of further reading provided at the end of the book is unhelpful and misleading.

But this caveat aside, the drawings and other illustrations, and Klotzko's narration, make the book highly approachable. Members of the public who would like to understand what the debate on human cloning is all about should read this book. ■

John Harris and Tuija Takala are at the Institute of Medicine, Law and Bioethics, University of Manchester, Manchester M13 9PL, UK.

Absolute beginnings

Degrees Kelvin: A Tale of Genius, Invention, and Tragedy

by David Lindley

Joseph Henry Press: 2004. 366 pp. \$27.95

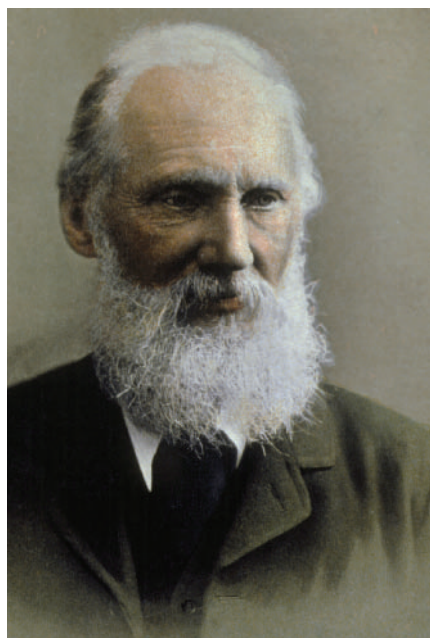
David B. Wilson

Recovering unfamiliar ideas from the past can be difficult, but David Lindley accepts the challenge, providing for the most part a fine survey of Lord Kelvin's life for a general audience. His narrative takes us from the child prodigy William Thomson to his later years as Lord Kelvin, the most honoured scientist and engineer of Victorian Britain. Lindley's early investigations formed the impression in his mind of a sharp contrast between the youthful Thomson, largely responsible for the science of thermodynamics, and the elderly Kelvin, unreasonably opposed to new scientific theories. Lindley's book explores that contrast.

Born in Belfast in 1824, Kelvin was only ten when he entered the University of Glasgow, where his father was the professor of mathematics. In his mid-teens, Kelvin published a paper correcting errors in a book by a professor of mathematics at the University of Edinburgh. From Glasgow, Kelvin went to the University of Cambridge, graduating with high honours in 1845. Extending work by the likes of Joseph Fourier and Michael Faraday, his early papers concerned heat, electricity, magnetism and light, as he presented mathematical analogies and mechanical models of these phenomena. In part, he was developing a theory of an elastic-solid ether, vibrations of which constituted light.

In 1846 Kelvin became professor of natural philosophy at the University of Glasgow. In the early 1850s, he and Rudolf Clausius established the modern science of thermodynamics. In the 1850s and 1860s, he was key to the successful laying of the Atlantic cable, which allowed telegraphic communication between the United States and Britain. In the mid-1860s, Kelvin used thermodynamics to refute the uniformitarian geology of the day, estimating that Earth was much younger than the geologists claimed, and he sought to unite ether and matter with a theory of smoke-ring-like vortex atoms.

Although his early research had inspired James Clerk Maxwell to develop his electromagnetic theory of light, Kelvin resisted Maxwell's conclusions and continued to seek suitable mechanical models of these phenomena. He also resisted darwinian evolution and aspects of the atomic models developed by J. J. Thomson and Ernest Rutherford to explain radioactivity. He became Lord Kelvin in 1892, resigned his professorship in 1899 and died in 1907.



Broad-minded? Lord Kelvin's work ranged from thermodynamics to electromagnetism.

Lindley's story of this remarkable life goes admirably beyond the ideas of science and engineering to reveal much of the day-to-day Kelvin. To this end, Lindley liberally quotes from a substantial family and scientific correspondence. As a Cambridge undergraduate, Kelvin received much fatherly advice about studying hard and behaving properly. The family's advice later turned towards Kelvin's preparation to apply for the Glasgow chair of natural philosophy when a vacancy seemed in the offing — his sister

advised Kelvin to grow a beard, to give him authority.

Other correspondence cited by Lindley reveals much of the relations between Kelvin and other scientists. In the early 1840s, Kelvin met G. G. Stokes, a Cambridge honours graduate of 1841, and their extensive correspondence attests to both their friendship and their common scientific interests, although the former seldom appeared explicitly in their letters, as Lindley notes. Another Cambridge honours graduate, P. G. Tait, championed Kelvin's physics, although their collaboration on the monumental *Treatise on Natural Philosophy* (1867) filled their letters with tales of Kelvin's delays and Tait's exasperation. Then there was longevity's sad side as colleagues died, Tait in 1901 and Stokes in 1903.

Lindley does a splendid job of explaining the scientific and technological concepts for a general audience, but at a deeper level the book falls short. His present-minded perspective too often misrepresents the historical context of ideas. Occasionally, subjects are misleadingly presented as not being well understood until the modern (that is, correct) theory appeared. The Michelson–Morley experiment of 1887 is mistakenly described as an attack on the ether, a view that is not really plausible without the hindsight provided by relativity theory some two decades later.

Applied to Kelvin's later dissents, this hindsight approach portrays him as being too practical and unimaginative to embrace new truths. However, such a judgment overlooks, for example, the significant scientific difficulties confronting natural selection — difficulties that caused Darwin to modify his views. Early atomic models faced similar problems, and acceptance of Maxwell's electromagnetic theory was hardly a straightforward matter. Indeed, in the late Victorian context, Kelvin's alternative to the current atomic models agreed with both his estimate of Earth's age and his objections to natural selection, forming a coherent and imaginative world view. Here, hindsight hides imagination.

If Lindley's extensive research insufficiently modified his early impression of Kelvin, his readers will nevertheless find much insight into the life and ideas of one of Victorian Britain's most interesting and influential men. It will certainly not replace *Energy and Empire* (Cambridge University Press, 1989) by Crosbie Smith and M. Norton Wise as the definitive biography of Kelvin, but at well under half the length it will undoubtedly be more accessible for most readers.

David B. Wilson is professor of history, mechanical engineering and philosophy at Iowa State University, Ames, Iowa 50011, USA. He is the author of *Kelvin and Stokes: A Comparative Study in Victorian Physics*.

New in paperback

Strange Matters: Undiscovered Ideas at the Frontiers of Space and Time

by Tom Siegfried

Berkley, \$15

"Not just a tour of current theoretical speculation on time, space and matter, but rather a cogent argument as to why we should take these exotic ideas seriously." Marc Kamionkowski *Nature* **420**, 362–363 (2002).

Rare Earth: Why Complex Life is Uncommon in the Universe

by Peter D. Ward & Donald Brownlee

Copernicus, £15.50, \$16.95

Small Worlds: The Dynamics of Networks between Order and Randomness

by Duncan J. Watts

Princeton University Press, \$24.95, £16.95

Spying with Maps: Surveillance Technologies and the Future of Privacy

by Mark Monmonier

University of Chicago Press, \$16, £11.50

The public cadaver

Anatomy: displays of bodies are no longer sufficient to explain the richness of modern anatomy to students or to the public.

Horst-Werner Korf and Helmut Wicht

Knowledge of anatomy underpins contemporary medicine, and is indispensable for understanding the structural and biological nature of humans. Dissection is used to analyse the body, and provide insights into its structure, function and dysfunction. Anatomists have to communicate their results to medical doctors and students — and to all those who are interested in their own body.

Anatomists have always had to walk the thin line between scientific objectivity and public spectacle. The first recorded dissections of human bodies, performed by Herophilus around 300 BC, were public, as were the few dissections carried out during the Middle Ages. In the seventeenth and eighteenth centuries, anatomical theatres were founded in many European cities. The dissections held in these theatres served partly to educate medical practitioners and students, but were also public spectacles of little educational value. In the nineteenth century, the rising numbers of students, the increasing amount of data, the intricacy and delicacy of the preparations, and finally, the use of microscopy requiring sophisticated histological equipment, forced anatomy to retreat into the well-equipped laboratories, dissection halls and lecture theatres of universities.

Modern anatomy consists not only of macroscopic examination of the body, but also views cells and molecules as essential players in any physiological or pathological process. It is no longer restricted to the static and descriptive levels, but has evolved into 'living anatomy'. Our anatomical knowledge has expanded tremendously, and today covers not only the spatial dimension of a dead body or cell, but also the temporal and dynamic dimensions underlying all living processes. For example, anatomical investigations of the circadian, rhythm-generating systems have located the internal clock in the hypothalamic suprachiasmatic nuclei, identified the neuronal pathway that serves the entrainment of the endogenous rhythm with the day–night rhythm, discovered photoreceptors in the retina responsible for this entrainment, and reached down to the level of the rhythmically expressed genes that make the clock tick. As a welcome side effect, anatomy remains one of the few medical disciplines that deals with the humans holistically.

Anatomy has come a long way from the 'show and tell' anatomy that could be taught



A closer look: but in autopsies today, dissection is only the start.

to a paying audience in an anatomical theatre within a few hours. Today, anatomy still requires the dissection of a human body and the display and naming of all its visible parts. But this macroscopic view is only the first part of the anatomical curriculum which embraces a 'vertical' holistic approach — from the macroscopical level (including self-examinations and demonstrations on living subjects) to the molecular (laboratory courses in molecular biology), from the spatial dimension (topographic anatomy) to the temporal (embryology, gerontology and evolutionary anatomy). This conceptual integration is central to the understanding and teaching of modern anatomy.

Those teaching anatomy, as for any other subject, need to capture the attention and interest of their pupils, but the engagement of an audience has to serve a genuine task — the transmission of knowledge. That is difficult enough with medical students, but the public also wants and needs to be informed about their bodies. Didactic principles cannot simply be borrowed from the spectacular anatomy of the old days — anatomy should be thrilling, but it should not offer cheap thrills. It is certainly not sufficient to display the macroscopy of the cadaver in public, to erect a cabinet of curiosities, to expand and explode and to plunder the history of art and anatomy, as Gunther von Hagens does in his *Body Worlds* shows. To demonstrate contemporary scientific anatomy to the public, it is necessary to combine contemporary concepts and cadavers, as the former are needed to understand the latter. This is a formidable task and requires sophisticated didactic methods and a commitment

from both teachers and the public that reaches beyond a brief visit to an exhibition.

Anatomy also has to respect a strict ethical code. Anatomical dissections occur within a diffuse and continuous transition zone between man and material. At the start there is the body of what was a self-determined living subject, and in the end, a dissected anatomical preparation, an object that can be used to obtain anatomical knowledge. But even though the integrity of a body is destroyed after death, the dignity of the person remains — thus both their corpse and their wishes deserve respect. Anatomical dissections or the removal of organs for scientific purposes should only be performed with the written consent of the donors. Sadly, this rule has been violated in the past and present. A recent example is the alleged sale of body parts donated to the Medical School of the University of California, Los Angeles (see *Nature* **428**, 243; 2004).

These problems underpin the necessity for non-commercial, professionally supervised 'willed body programmes' with high transparency. For example, our anatomy school will not accept any dead body unless the deceased gave prior written consent. In turn, we treat the corpse with due respect: we use the cadaver only for scientific and educational purposes and we provide a decent funeral of the donor's remains if this is his or her wish.

We need anatomy, but on the basis of contemporary concepts, methods, didactics and ethics. Medicine needs anatomy. The public needs anatomy. Anatomy needs the public, as an audience and as a donor. What we need least is the public cadaver exposed in shows of little scientific and educational value. ■

Horst-Werner Korf and Helmut Wicht are in the Senckenbergische Anatomie, Medical Faculty of the Johann Wolfgang Goethe-University, Frankfurt/Main, Theodor-Stern-Kai 7, Frankfurt D-60590, Germany.

FURTHER READING

Haeser, H. *Lehrbuch der Geschichte der Medizin* Vol. 1. (Friedrich Mauke, Jena, 1853).

Korf, H.-W. & Stehle, J. H. (eds.) *Cell Tissue Res.* **309**, 1–199 (2002).

Stukenbrock, K. *"Der zerstückte Körper". Zur Sozialgeschichte der anatomischen Sektionen in der frühen Neuzeit (1650–1800)* (Steiner, Stuttgart, 2001).

Yamaguchi, S. et al. *Nature* **409**, 684 (2001).

Tall storeys

Ian Woodward

The tallest living organisms are trees, but how tall could they be and what stops them growing any taller? Measurements at the tops of the world's tallest trees now provide quantitative answers to these questions.

Record-holders are of enduring interest¹, and in the plant world none stands out more than the tallest tree on Earth. This tree, coincidentally the tallest living organism, is found in California. It is a giant redwood (*Sequoia sempervirens*), which, at 112.7 m high, is equivalent to a 30-storey building.

Theories about the control of tree height have moved from a realization that trees are more than sufficiently engineered to cope with the mechanical stresses of height, to focus on their capacity to deliver water reliably to the treetop^{2,3}. However, until now there has been no direct test of this at the top of the tallest trees. In a breathtaking field study described on page 851 of this issue⁴, Koch and colleagues have now done so, carrying out research on five of the eight tallest trees in the world (all giant redwoods), including the tallest (Fig. 1) and second tallest. The results of linked field and laboratory studies indicate that the fundamental control of maximum tree height is water supply to the treetop.

Water rises to treetops by transpiration, the process by which water evaporates from leaves through stomata, the pores on the leaf surface. Transpiration pulls water into the plant through the roots and up to the very top through the water-conducting cells of the xylem tissue. This pull overcomes the opposing forces of gravity and friction, with the greatest pull occurring at the treetop. Water columns endure this tension until they reach a threshold at which they break, undergoing embolism with the introduction of air bubbles. Koch *et al.* measured the maximum tension on the water column above 100 m in the tallest redwoods. They found that the maximum tension is close to the point of embolism, establishing this value as the first major control on height.

Water reaching the leaves at the top of the tree actively drives cellular growth. Growth occurs by pressurizing the cells with a flow of water from the xylem, after the gravitational and frictional forces at the treetop have been overcome. These forces increase monotonically with height, decreasing the growth capacity of water flow^{4,5}.



Figure 1 High and mighty. This composite image of trees in the Humboldt Redwoods State Park includes the 'Stratospheric Giant', the tallest tree on Earth. The image consists of more than 700 photos taken by James Balog — spot the human climbers!

One effect of the reduced water flow is that cells at the top of the tree are small, with thick cell walls, and consequently the leaves produced are small and dense. At 110 m, redwood leaf density is the highest ever recorded, suggesting that growth has been significantly impaired: this is a second limitation on height. The increased invest-

ment in non-photosynthetic leaf tissue leads to a constant photosynthetic decline per unit of tissue investment to the treetops⁴; this fall in photosynthetic efficiency imposes a third constraint on height.

Maintaining xylem water status above the point of embolism affects gas exchange in the leaf, where carbon dioxide is absorbed for photosynthesis and oxygen is produced as a waste product. Koch and colleagues' *in situ* measurement of photosynthesis at over 110 m, a remarkable achievement in itself, shows the leaf CO₂ concentration to be close to the lowest ever measured at ambient atmospheric CO₂ concentrations. This limitation on CO₂ diffusion through the stomata imposes a fourth control on height.

The four physiological constraints on tree height are shown in Fig. 2 (overleaf), from which a narrow range (122–130 m) for the maximum possible tree height can be predicted. If this is correct, then the tallest trees, which may be more than 2,000 years old, still have some way to go. The observation that they are continuing to grow at about 0.25 m per year supports that idea.

These magnificent redwoods offer features that only the tallest of trees can deliver; the work of Koch *et al.* is only the latest in a line of very diverse, but highly relevant research on tall trees. For example, flow through the xylem is slow⁶: water entering the base of a redwood trunk could take as long as 24 days to reach the treetop. During periods of drought, this long delay in the replenishment of transpired water could lead to closure of stomata and inhibition of photosynthesis. But tall trees have a water store in the form of the water-conducting sapwood — an outer zone of wood, only 4–5 cm deep, that lies between the heartwood and the bark. In 60-m Douglas fir⁷ trees, this stored water provides 18% of the water supply for photosynthesis, nearly double that observed for a 15-m tree. The sapwood may be only 2–3% of the total trunk volume, but in the tallest trees it constitutes a substantial liquid reservoir.

Tall trees use considerable quantities of

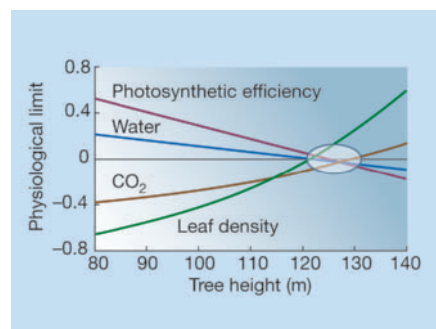


Figure 2 Physiological responses to tree height². Each curve is normalized so that a zero value indicates the maximum sustainable tree height based on photosynthetic efficiency, the point of xylem embolism (water), CO₂ supply to the leaf, and leaf density. The ellipse shows the predicted range of maximum heights. Water supply is the key constraint on height, given that it is essential for the other physiological factors.

water. For example, a 45-m redwood uses about 600 kg of water each day⁸, a figure that increases substantially with height and size⁷. It seems surprising, therefore, that the redwoods live in a climate with an annual dry season of 3–4 months. Offsetting such an apparent drawback, however, is the oceanic influence on local climate, which means that dry-season fog occurs for up to two weeks at a time: fog reduces transpiration, a benefit in the dry season. Moreover, tall trees actually increase the interception and capture of fog coming in off the sea, to the tune of 34% of the annual incidence of precipitation; in

their absence, the precipitation input from fog is halved.

Koch *et al.*⁴ have taken research on redwoods to new heights in testing and quantifying theories about the control of tree growth. There is substantial scope for further research, however, not least in exploring the processes that lead to the particular limiting values used to predict maximum height, as depicted in Fig. 2. Finally, it is remarkable how research by Koch's group has evolved: this latest paper dealing with redwoods follows early research on the lowly radish⁹. Yet that is no real surprise: despite the very different packaging and longevity of the two species, their physiological processes are much the same.

Ian Woodward is in the Department of Animal and Plant Sciences, and the NERC Centre for Terrestrial Carbon Dynamics, University of Sheffield, Sheffield S10 2TN, UK.

e-mail: f.i.woodward@sheffield.ac.uk

1. Folkard, C. *Guinness World Records 2004* (Guinness World Records, London, 2004).
2. Sperry, J. S., Hacke, U. G., Oren, R. & Comstock, J. P. *Plant Cell Environ.* **25**, 251–263 (2002).
3. Hacke, U. G. & Sperry, J. S. *Plant Cell Environ.* **26**, 303–311 (2003).
4. Koch, G. W., Sillett, S. C., Jennings, G. M. & Davis, S. D. *Nature* **428**, 851–854 (2004).
5. Woodruff, D. R., Bond, B. J. & Meinzer, F. C. *Plant Cell Environ.* **27**, 229–236 (2004).
6. Sala, A., Carey, E. V. & Callaway, R. M. *Oecologia* **126**, 42–52 (2001).
7. Phillips, N. G. *et al.* *Tree Physiol.* **23**, 237–245 (2003).
8. Dawson, T. E. *Oecologia* **117**, 476–485 (1998).
9. Koch, G. W., Schulze, E. D., Percival, F., Mooney, H. A. & Chu, C. *Plant Cell Environ.* **11**, 755–767 (1988).

Applied physics

Speed limit ahead

C. H. Back and D. Pescia

Are there any limits to what science and technology can achieve? When it comes to recording data in magnetic media, the answer is yes: there is a natural limit to the speed at which data can be encoded.

In magnetic recording, information 'bits' are stored in regions of a magnetic material as opposite magnetizations or spins, corresponding to the values '1' and '0'. High information density is a major requirement, but even the highest density is useless if the information cannot be stored and retrieved sufficiently quickly. On page 831 of this issue, Tudosa *et al.*¹ show that there is likely to be a limit to how quickly a bit of magnetic information can be reliably written. This minimum time is on the picosecond scale — 10⁻¹² seconds — which is a thousand times faster than has so far been achieved in state-of-the-art devices. However, this is still ten thousand times slower than the duration of the shortest laser pulses that are already available² and would ideally be used for data-writing.

The standard way to reverse the direction of a spin, and thereby write or re-write information, is to apply a magnetic field along the spin axis. This is a 'zero-torque' configuration: the torque is the vector product of the spin and magnetic field and hence is zero when these are aligned. Fluctuations such as thermal excitations can, however, create a misalignment between the spin and the applied magnetic field, generating a non-zero torque that will cause the spin to change its direction. This 'viscous switching' has been the method of choice because the rise times of magnetic-field pulses were limited, but it is a rather slow process, of some nanoseconds' duration.

However, a basic knowledge of magnetic resonance suggests that reversing the spin requires a magnetic-field pulse that is per-

pendicular, not parallel, to the spin direction. This is a maximum-torque configuration. The spin changes its direction by precessing (or turning) around the applied field direction, doing so at maximum speed³ as defined by the Larmor frequency of precession. Unfortunately there is a drawback to precessional switching: once the spin begins to precess, it does not necessarily stop when it reaches the required, opposite direction⁴.

Although this problem might be overcome³, a new obstacle in the path of viable precessional-switching technology has been revealed by Tudosa and colleagues¹. Their experiment is the first to use the magnetic-field pulse generated by a tightly packed bunch of electrons from the Stanford Linear Accelerator in California as it passes through the centre of a magnetic recording film at almost the speed of light. This magnetic-field pulse is probably the shortest and most powerful available on Earth, with a maximum strength of about 3 tesla and a duration of only 2.3 picoseconds.

The pulse causes switching, provided the minimum angle of precession ($\pi/2$) is achieved. This means that if the pulse duration is reduced, the magnetic-field strength must be increased so that the product of duration and strength is still larger than the critical value needed for switching to occur. The critical value depends on the initial direction of each spin just before and during application of the magnetic-field pulse. For smooth recording, it is essential that the initial state of every spin is well defined and uniform while the pulse is in place. Otherwise some spins will switch, others will relax back to the initial state and the final state of the switching operation will no longer be a well-defined bit.

In their experiment, Tudosa *et al.*¹ detected switching through the pattern of magnetic domains left behind after exposure to successive electron bunches. The field strength decays as the inverse of the distance from the bunch, so although some regions of the magnetic film will undergo one or more switching processes, other more distant regions will remain unaffected. A pattern of black and white concentric rings develops, corresponding to domains of opposite magnetization (see Fig. 1 on page 832).

If precessional switching were totally reliable, repeated field pulses would cause the domains to change, from black to white and back again, in a very precise way. But Tudosa *et al.* report that, although the overall picture does indeed reflect black-white domain interchange, there is a remarkable deterioration in the sharpness of the boundaries separating the domains during repeated pulsing. This deterioration matches the predictions of a very simple model that assumes some small non-uniformity in the initial spin directions. Such non-uniformity affects the switching field and introduces a kind

of chaos into a system that should ideally perform in exactly the same way over many writing cycles.

The origin of the non-uniform initial state is still a mystery, although the authors might be on the right track when they suggest that the very strong magnetic fields generated by the electron pulse excite non-uniform precessional modes. Because such dangerous 'deformations' of the spin distribution must be avoided to achieve reliable recording, there is an intrinsic limit on the minimum switching time between magnetization states. No matter how short and strong the magnetic-field pulse, magnetic recording cannot be made ever faster. Tudosa and colleagues' finding and their

interpretation of it are bound to trigger numerous follow-up experiments, such as measuring the switching trajectory on a microscopic scale.

C. H. Back is at the Institut für Experimentelle und Angewandte Physik, University of Regensburg, Universitätsstrasse 31, 93040 Regensburg, Germany.

e-mail: christian.back@physik.uni-regensburg.de

D. Pescia is at the Laboratorium für Festkörperphysik, ETH Zürich, CH-8093 Zürich, Switzerland.

e-mail: pescia@solid.phys.ethz.ch

1. Tudosa, I. *et al.* *Nature* **428**, 831–833 (2004).

2. Kienberger, R. *et al.* *Nature* **427**, 817–819 (2004).

3. Gerrits, Th., van den Berg, H. A. M., Hohlfield, J., Bär, L. & Rasing, Th. *Nature* **418**, 509–511 (2002).

4. Acremann, Y., Back, C. H., Buess, M., Pescia, D. & Pokrovsky, V. *Appl. Phys. Lett.* **79**, 2228–2230 (2001).

Genomic imprinting

Mice without a father

David A. F. Loebel and Patrick P. L. Tam

In mammals, genomes from both parents are generally needed to make viable offspring. But changing the expression of 'imprinted' genes can render the father's contribution dispensable.

Sexual reproduction in animals ensures that each individual normally inherits one set of genes from each parent. But viable offspring that have only maternal genes — none from the father — can be produced through parthenogenetic reproduction in plants and most groups of animals. A notable exception is mammals, in which fathers remain a necessity. But why is this so? On page 860 of this issue, Kono and colleagues¹ provide the most compelling evidence so far that a phenomenon called genomic imprinting is the stumbling block.

The importance to mammals of a proper combination of parental genes is highlighted by studies of mouse embryos that contain only maternally or paternally derived chromosomes. In embryos containing only female-derived chromosomes, the 'extra-embryonic' tissues that are required to support embryonic growth develop poorly, and the embryo dies soon after it has implanted into the womb^{2,3}. By contrast, the development of embryos that contain only paternal chromosomes is retarded, but the extra-embryonic tissues develop comparatively well^{2,3}. This seems to indicate that the paternal copies of some genes are more important for development of the extraembryonic tissues, and that maternal copies of the others are more essential for fetal development.

The most likely explanation for this is that the maternal and paternal genomes are not exactly equivalent, but are endowed with different 'imprints', which lead to differential gene expression in the embryo. Imprints, or 'epigenetic modifications', mark the two copies of each gene as being inherited either

from the mother or from the father. They are chemical changes to DNA or to chromosomal proteins that are heritable through cell divisions, but do not involve changes to the DNA sequence.

Imprinted regions of the genome typically cover large chromosomal domains of 1 million base pairs or more, and may involve the coordinated regulation of several genes. *Igf2* and *H19*, for instance, are imprinted genes that are located on mouse chromosome 7, 100 kilobases apart. These two genes are oppositely imprinted — expression is from the maternal copy of *H19* and the paternal copy of *Igf2* (Fig. 1, overleaf). The regulation of expression involves several nearby stretches of DNA, which act as enhancers, promoters and an insulator of gene expression.

The insulator (or boundary element) is upstream of *H19*, between this gene and *Igf2*, and is identified as a differentially methylated domain (DMD). According to a current model, this region cooperates with enhancer elements that are downstream of *H19*. On the maternal chromosome, an enhancer-blocking protein (CTCF, which recognizes the DNA sequence CCCTC) binds to the DMD, preventing the enhancers from interacting with the promoter of *Igf2*, and instead favouring *H19* expression^{4,5}. But on the paternal chromosome, the DNA of the boundary element is methylated; the blocking protein cannot bind, and *Igf2* can be expressed while *H19* is not.

Epigenetic modifications of the maternal genome occur during oocyte (egg) maturation. In earlier work, Kono and colleagues⁶



100 YEARS AGO

A Study of British Genius. Mr. Havelock Ellis recognises three great foci of intellectual ability in England:— (1) the East Anglian focus; (2) the south-western focus; and (3) the focus of the Welsh Border. The first of these is the most recent and the most mixed ethnologically, as East Anglia is very open to invasion, and all kinds of foreigners have settled there. The second is the largest and oldest, and the population has much darker hair; it may be called the Goidelic-Iberian district. The district is defended by Wansdyke and Bokerley Dyke. The third is termed the Anglo-Brythonic district. The Anglo-Danish part of England — Lincolnshire, Nottinghamshire, Derbyshire, Yorkshire, and thence into Scotland — has its own peculiar anthropological characters. Its children have usually been more remarkable for force of character than for force of intellect. East Anglia is productive of great statesmen, ecclesiastics and scholars, and of musical composers and painters. It has no aptitude for abstract thinking; its special characters seem to be humanity, patience, grasp of detail, and love of liberty. The people of the south-western focus are sailors rather than scholars, and courtiers rather than statesmen; they are innovators and pioneers in the physical and intellectual worlds, and, above all, are impressive, accomplished, and irresistible personalities. The genius of the Welsh Border is artistic in the widest sense, and notably poetic; there is a tendency to literary and oratorical eloquence, frequently tinged with religious or moral emotion, and there are no scientific men of the first order. From *Nature* 21 April 1904.

50 YEARS AGO

The announcement in *The Times* of April 12 of the production of element number 100 by Prof. G. T. Seaborg and his collaborators at the University of California follows closely on their discovery of element 99... Identification of the new isotope was presumably based on the α -decay systematics, built up by Prof. Seaborg and others, which have proved extremely reliable in this field. Element 100 is stated to behave chemically like erbium, its analogue in the rare-earth group. There is no reason for believing that this will be the last new element to be prepared, although the increasing probability of spontaneous fission as the atomic number advances would appear to limit the total number of elements to about 110. From *Nature* 24 April 1954.

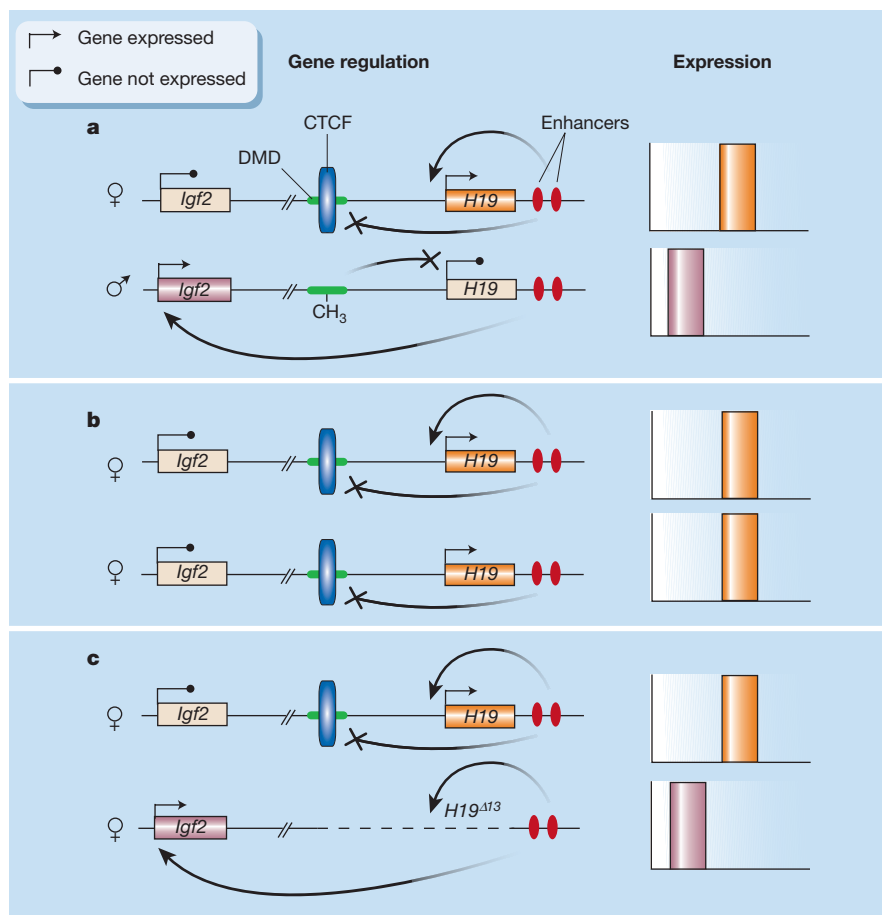


Figure 1 No need for males? **a**, The *H19* and *Igf2* genes are found on the same mouse chromosome and are oppositely 'imprinted': in normal embryos, *H19* is expressed only from the maternal chromosome and *Igf2* only from the paternal chromosome. On the maternal chromosome, the protein CTCF binds to the differentially methylated domain (DMD), blocking the access of enhancers to *Igf2*. Thus, *H19* is expressed instead. On the paternal chromosome, the DMD is methylated (represented by CH₃), and so CTCF cannot bind; the enhancers therefore have access to *Igf2*, which is expressed, whereas *H19* is silenced. Modified from ref. 5. **b**, In conventional parthenotes, which have two maternal genomes, no *Igf2* is expressed and the embryos die. **c**, Kono *et al.*¹ combined the chromosomes from a fully grown egg (which has all maternal imprints) with the chromosomes from a non-growing egg from which the DMD and *H19* were deleted (*H19*^{Δ13}). These deletions mimicked the absence of paternal *H19* activity and enabled *Igf2* expression, leading to viable adults.

found that by combining the chromosomal complements of a non-growing egg from a newborn mouse and a mature, fully grown egg, parthenogenetic embryos could survive for longer than normal. Because one set of chromosomes came from an earlier stage of egg development, it had not acquired the maternal imprints and so might have been able to express some genes that normally would be expressed only from the paternal chromosome.

To see whether further altering genomic imprinting might influence the developmental potential of parthenogenetic embryos, Kono and colleagues⁷ used the same techniques to test whether the 'unimprinted' genome of a non-growing egg that lacked the *H19* gene region⁸ had any impact on embryonic viability. Indeed, parthenogenetic embryos that lacked one copy of *H19* survived nearly to term, but died with a poorly formed placenta. These embryos evidently

lacked *H19* expression from one set of chromosomes — as do normal embryos. But because the DMD remained present and unmethylated on the chromosome from the non-growing egg, it presumably prevented *Igf2* expression.

Kono and colleagues¹ have now carried out similar experiments with mice in which both *H19* and the DMD are deleted⁹; they refer to the animals as *H19*^{Δ13} mice (Fig. 1). The authors combined chromosomes from non-growing oocytes from these mutants with chromosomes from mature oocytes of normal mice. In theory the deletion of *H19* from the non-growing oocyte's genome should mimic the lack of paternal *H19* activity, and, because the DMD is missing too, the blocking protein cannot bind to it; the upshot should be that *Igf2* can be expressed, as would be paternally derived *Igf2*. Even if we assume all this, however, the result is surprising: two apparently normal, live female

pups were born. One of them reached adulthood, and even produced offspring.

A further surprise came from microarray analysis of the expression of over 11,000 genes. The expression levels of more than 1,000 genes in the surviving *H19*^{Δ13}-carrying parthenotes were more similar to those of normally fertilized embryos than to those of parthenotes with two intact copies of *H19*. This widespread effect on gene expression cannot be put down to a direct effect of the lack of one copy of *H19*. One possibility is that the changes in gene expression are indirect effects of the improved growth of the *H19*^{Δ13} parthenotes. But it's amazing that altering the expression of just two imprinted genes can have a ripple effect on the rest of the genome.

Of particular interest is the effect on the expression of other imprinted genes. The expression of *H19* and *Igf2*, as well as of *Dlk1* and *Gtl2* — another pair of oppositely imprinted genes, which are located on a different chromosome to *H19* and *Igf2* — in normally developing *H19*^{Δ13} parthenotes was comparable to that of normal mice. But in growth-retarded *H19*^{Δ13} parthenotes, *Dlk1* and *Gtl2* were not expressed at normal levels. Taken as a whole, the work of Kono *et al.*^{1,6,7} provides good evidence that incorrect expression of imprinted genes is one of the major reasons why natural parthenogenesis in mammals has not been possible. What is still not understood is why such a barrier to single-parent reproduction has evolved.

This work¹ also raises the question of why only a small proportion of the embryos survived, and why the effects on gene expression are so variable. Perhaps some of this can be explained by the outbred nature of the mouse strains used. It will be particularly pertinent to determine how altering the activity of the *H19* gene can change the expression of so many others, especially that of apparently unconnected imprinted genes. Until we fully understand the role and regulation of imprinted genes in development, it seems that the participation of the father in reproduction will remain necessary.

David A. F. Loebel and Patrick P. L. Tam are in the Embryology Unit, Children's Medical Research Institute, University of Sydney, Wentworthville, New South Wales 2145, Australia.
e-mails: dloebel@cmri.usyd.edu.au
ptam@cmri.usyd.edu.au

1. Kono, T. *et al.* *Nature* **428**, 860–864 (2004).
2. Surani, M. A., Baron, S. C. & Norris, M. L. *Nature* **308**, 548–550 (1984).
3. Barton, S. C., Surani, M. A. & Norris, M. L. *Nature* **311**, 374–376 (1984).
4. Thorvaldsen, J. L., Duran, K. L. & Bartolomei, M. S. *Genes Dev.* **12**, 3693–3702 (1998).
5. Arney, K. L. *Trends Genet.* **19**, 17–23 (2003).
6. Kono, T., Obata, Y., Yoshimizu, T., Nakahara, T. & Carroll, J. *Nature Genet.* **13**, 91–94 (1996).
7. Kono, T., Sotomaru, Y., Katsuzawa, Y. & Dandolo, L. *Dev. Biol.* **243**, 294–300 (2002).
8. Ripoché, M. A., Kress, C., Poirier, F. & Dandolo, L. *Genes Dev.* **11**, 1596–1604 (1997).
9. Leighton, P. A., Ingram, R. S., Eggenschwiler, J., Efstratiadis, A. & Tilghman, S. M. *Nature* **375**, 34–39 (1995).

The mantle deformed

Sébastien Merkel

What happens to minerals under the conditions characteristic of the Earth at great depths? Experiments performed under such conditions illustrate how the main constituent of the lower mantle may behave.

Until the middle of the twentieth century, it was thought that the Earth is a solid body and that no internal movement of matter could occur. Then, with the discovery of plate tectonics and mantle convection, a whole new understanding of the planet's dynamics emerged. Matter can indeed flow within the seemingly solid regions of the deep Earth, and these processes often control the surface dynamics. Yet the microscopic mechanisms controlling flow remain poorly understood. That is scarcely surprising: it is no easy matter to simulate the conditions to which materials in, for example, the lower mantle or the inner core are subject. So the plastic properties of those materials have remained largely unknown.

On page 837 of this issue, however, Cordier *et al.*¹ describe new results on the plastic properties of the main constituent of the lower mantle, a mineral known as silicate perovskite (Fig. 1). They have exploited both experimental and theoretical advances to demonstrate that dislocations in silicate perovskite can be activated under the conditions of the lower mantle. This implies that silicate perovskite could develop large-scale anisotropic structures — that is, with physical properties that depend on the direction of observation — and that it may be possible to measure these structures using seismological methods.

Materials inside the solid layers of the Earth are not perfect crystals but rather aggregates of crystalline grains — polycrystals — that are undergoing plastic deformation. The process is governed by the interaction between deformation within individual grains (through the movement of dislocations), sliding and diffusion along grain boundaries, and the stress and strain fields applied by the environment. Dislocations are line defects within the crystal structure and their motion in the lattice causes slip and plastic deformation within the grain. Depending on external conditions such as pressure and temperature, and the physical properties of the crystal, slip is observed to occur in specific planes and directions, known as slip systems.

If dislocations dominate the deformation, grains will deform preferentially in certain directions and the polycrystal will develop anisotropy in such physical properties as elasticity and electrical conductivity. In the Earth, the anisotropic elastic proper-

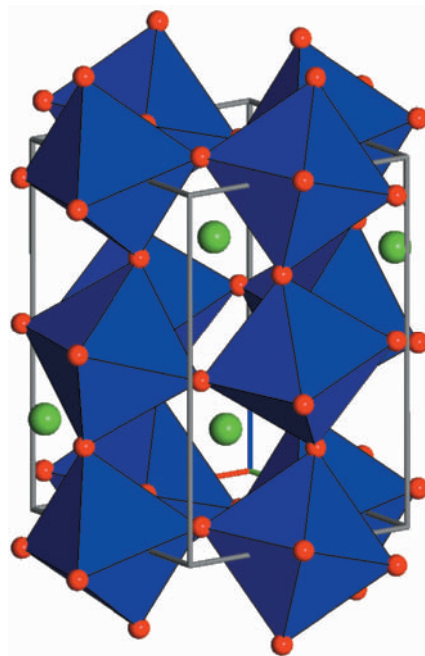


Figure 1 The structure of silicate perovskite (Mg,Fe)SiO₃. Red circles represent oxygen atoms and green circles magnesium or iron. The silicon atoms are located at the centre of the blue SiO₆ octahedra. Cordier *et al.*¹ add new data on how this material deforms under the conditions of the Earth's lower mantle, and so on anisotropic structures that might be evident with seismological methods.

ties can be studied using seismic techniques, which measure the directional dependency of seismic velocities. Thereafter, a combination of rheological and seismological measurements can be used to interpret the dynamic geological processes deep inside the Earth².

The transformation of mantle minerals into silicate perovskite (Mg,Fe)SiO₃ and magnesiowüstite (Mg,Fe)O at a pressure of about 23 gigapascals marks the limit between the upper and lower mantle at a depth of 660 km. On the evidence of high-pressure experiments, seismological measurements, geochemistry and numerical modelling, perovskite is believed to account for about 80% of the lower mantle. Understanding its properties is therefore essential to understanding the physical and geological processes that occur within this layer. Unfortunately, silicate perovskite is unstable under low pressure and high temperature, and it is also highly sensitive to the electron

irradiations used in microscopy. So silicate perovskite cannot be studied with the techniques used to investigate the plastic properties of crustal or upper-mantle minerals, and the microscopic processes controlling its plasticity remain unknown.

To address this issue, experiments have been carried out on analogues that have the perovskite structure. These studies led to the conclusion that, under lower-mantle conditions, diffusive processes should dominate perovskite deformation. This in turn implied that silicate perovskite could not generate anisotropic structures in the lower mantle³. But it has also been shown that deducing the plastic properties of perovskite-structured materials from analogues can be misleading. So the analogue approach has to be treated with caution⁴, and direct investigations of silicate perovskite are necessary.

In recent years, techniques have been developed that allow the experimental deformation of minerals under pressures and temperatures close to those of the lower mantle. But analysing the properties of the sample to identify the deformation mechanisms involved is a further challenge. Cordier *et al.* have overcome that challenge by adapting an approach — X-ray peak broadening analysis — developed in metallurgy. As in electron microscopy, where dislocation contrast is strong or disappears depending on certain conditions, X-ray diffraction peaks become broader or narrower. These peaks can be used to characterize dislocations within the material.

Taking full advantage of the new techniques, Cordier *et al.*¹ have produced the first experimental evidence that dislocations can indeed be activated in silicate perovskite under lower-mantle conditions. They were also able to identify the most active slip systems — that is, the plane and direction in which the slip occurs. Key questions, such as the influence of grain size, strain rate or phase mixing, remain to be solved. Nevertheless, these new results force us to reconsider our thinking about lower-mantle rheology. Seismological observations identify highly anisotropic patches at the bottom of the Earth's lower mantle, whereas the bulk of the deep mantle seems largely isotropic^{5–7}. But numerical modelling indicates that other regions could also exhibit anisotropy, for instance in the vicinity of subducting slabs where the Earth's crust is driven back into the mantle⁸.

For many years, investigations of the dynamics of the Earth's lower mantle were neglected because of the lack of appropriate seismological and rheological data. Before too long, however, the new interdisciplinary interactions between areas such as high-pressure mineralogy, studies of mineral plasticity, engineering, seismology and geodynamics look set to revolutionize

understanding of our planet's deeper layers. ■

Sébastien Merkel is at the Institute for Solid State Physics, University of Tokyo, Kashiwanoha 5-1-5, Kashiwa, Chiba, 277-8581 Japan.

e-mail: smerkel@issp.u-tokyo.ac.jp

1. Cordier, P., Ungár, T., Zsoldos, L. & Tichy, G. *Nature* **428**, 837–840 (2004).
2. Wenk, H. R. *Rev. Mineral. Geochem.* **51**, 291–330 (2002).
3. Karato, S., Zhang, S. & Wenk, H. R. *Science* **270**, 458–461 (1995).
4. Beauchesne, S. & Poirier, J. P. *Phys. Earth Planet. Inter.* **61**, 182–198 (1990).
5. Meade, C., Silver, P. G. & Kanshima, S. *Geophys. Res. Lett.* **22**, 1293–1296 (1995).
6. Lay, T., Williams, Q. & Garnero, E. J. *Nature* **392**, 461–468 (1998).
7. Panning, M. & Romanowicz, B. *Science* **303**, 351–353 (2004).
8. McNamara, A. K., van Keken, P. E. & Karato, S. *Nature* **416**, 310–314 (2002).

Cancer

Enzymes play molecular tag

Deborah K. Morrison

The B-RAF protein is often mutated in human cancers, contributing to their development. Although most known mutations stimulate its catalytic activity, others, surprisingly, impair it — yet still cause cancer.

Like most other cellular events, cell proliferation is tightly regulated by signals from the surrounding environment. These cues are relayed from the cell surface to the nucleus by defined signal-transduction cascades. The core components of one such pathway are the RAS, RAF, MEK and ERK proteins. If this pathway is constantly

switched on, it can cause cells to proliferate wantonly, resulting in cancer.

Researchers have known for some time that there are cancer-promoting mutations in the RAS protein that keep it permanently 'on'. Recently, large-scale genomic screens have also detected mutations in one member of the RAF family of proteins — B-RAF —

in 65% of malignant melanomas¹ and many colorectal², ovarian³ and papillary thyroid^{4,5} cancers. How these mutations alter B-RAF's function is the topic of an elegant study, published in *Cell*, by Wan and colleagues⁶. The paper also provides structural information that should help to guide the search for more effective inhibitors of this protein family.

The RAF enzymes are central intermediates in this fundamental signalling cascade, transmitting signals from RAS to the downstream enzymes MEK and ERK (Fig. 1a, overleaf). In mammalian cells there are three members of the RAF family, A-RAF, B-RAF and C-RAF. Working out how these proteins are regulated has been a daunting task, largely because of the complexity of the process. There seem to be several mechanisms, including self-inhibition (involving a regulatory domain located at one end, the amino terminus, of the proteins), interactions with binding partners such as RAS and the 14-3-3 protein, and the phosphorylation of (covalent linkage of phosphate groups to) inhibitory and activating sites on the proteins^{7,8}.

The activation of RAF is typically

Developmental genetics

Bittersweet evolution

Structures that occur in closely related organisms and that look the same are usually considered to be homologous — their similarity is taken to arise from their common ancestry. Common sense suggests that the more complex such structures are, the less likely they are to have evolved independently and the more valuable they should be for studying systematics. But what if 'obviously' identical organs have arisen through two mutually exclusive developmental routes?

Beverly Glover and colleagues have revealed one such case (*Gene* **331**, 1–7; 2004). It occurs in the floral organs of the genus *Solanum* from the nightshade family. In one group of these species, the anthers — the flower's pollen-producing organs — are arranged as a cone, which functions like a 'pepperpot' (see the yellow, cone-like structures in a and b, right). In the pepperpot of bittersweet (*S. dulcamara*), the anther surfaces are held together by a glue-like secretion (a). In another species from the same group, tomato (*S. lycopersicum*), they are instead linked by interlocking hairs, or

trichomes, along the edges of the anthers (b).

Glover *et al.* find that tomato trichomes are clearly required for pepperpot formation. In one form, the *dialytic* mutant, which lacks them, the pepperpot fails to develop (d). In bittersweet, however, trichomes surprisingly prevent pepperpot formation. Glover *et al.* show this using transgenic plants in which expression of a gene from snapdragon leads to the development of hairs on bittersweet anthers. The hairs push the glue-bearing surfaces apart, preventing pepperpot formation (c).

This result makes it unlikely that the tomato-type pepperpot originated from the bittersweet type, or vice versa, because the development of anther hairs in bittersweet-type cones would probably have caused the cone to fall apart, whereas the addition of glue to tomato-type cones already supported by trichomes would probably have carried no selective advantage. So the most plausible conclusion is that pepperpots originated twice independently in the lineages that led to tomato and bittersweet.

Molecular systematic analysis



confirms that tomato and bittersweet are closely related, and the traditional view would be that their pepperpot cones are obviously homologous. But genetic tinkering and mutant analysis show that they probably are not — that they are convergent, having taken different routes to the same end. Life's potential to invent complex

structures more than once may worry systematists, who depend on reliable characters to reconstruct relationships between organisms. But it will please anyone who admires nature's innovative power.

Günter Theißen

Günter Theißen is at Friedrich Schiller University, D-07743 Jena, Germany.
e-mail: guenter.theissen@uni-jena.de

A. DAVIS, JOHN INNES CENTRE

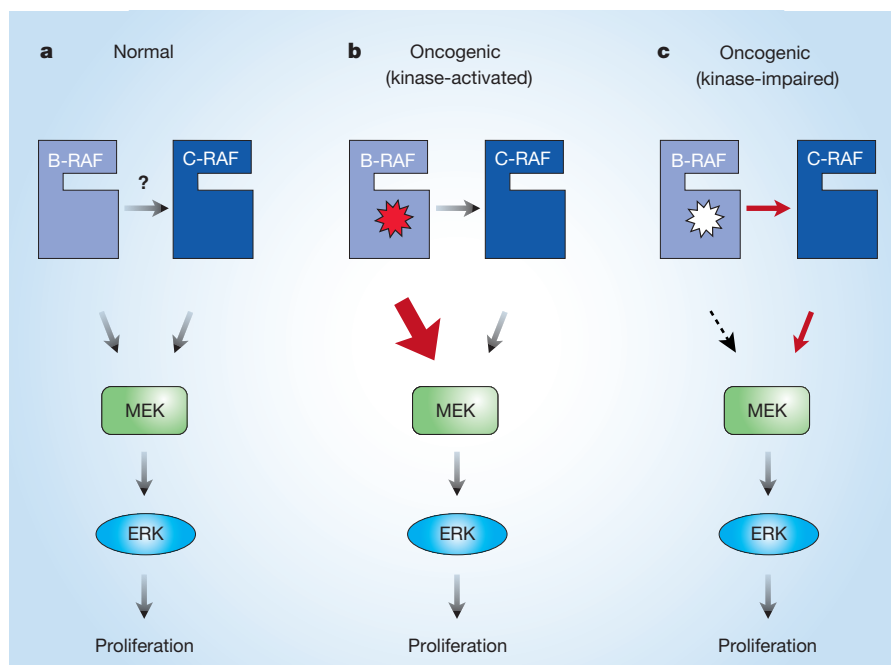


Figure 1 Cells, signalling cascades and cancer. **a**, In one of the best-characterized signalling pathways that promotes cell proliferation, signals from outside the cell are passed via the RAS protein (not shown) to activate B-RAF and C-RAF, both of which activate MEK, which in turn switches on ERK. **b**, Most cancer-promoting (oncogenic) mutations in B-RAF stimulate its catalytic activity, resulting in MEK and ERK activation without the earlier events. **c**, Other oncogenic mutations impair B-RAF's kinase activity but somehow allow it to activate C-RAF, which stimulates MEK and ERK in its place. Although kinase-activated B-RAF mutants can also activate C-RAF, this is not required for later events. Whether B-RAF activates C-RAF under normal conditions is unknown.

initiated by an interaction with RAS, which leads to RAF relocating from the cellular cytoplasm to the plasma membrane. This interaction also promotes conformational changes in RAF that remove self-inhibition and enable phosphorylation of its activating sites. In the mammalian C-RAF and A-RAF proteins, these sites are located in two regions of the catalytic ('kinase') domain — namely the negatively charged regulatory region (the N-region) and the activation segment^{9,10}.

In contrast, the N-region of B-RAF is constantly phosphorylated, so phosphorylation of the activation segment alone is enough to turn the enzyme on. Because this continuous phosphorylation of the N-region also disrupts self-inhibition, B-RAF is the member of this family that teeters closest to the brink of activation. This may explain why it is the only RAF protein so far that has been found to be activated by mutation in human cancers.

Of the more than 30 known tumour-associated B-RAF mutations, around 90% alter amino acids in the kinase domain, raising the question of whether these mutations promote cancer simply by increasing B-RAF's catalytic activity (Fig. 1b). To find out, Wan and colleagues⁶ characterized the enzymatic properties of 22 such mutants. They found that most can be explained in this manner, including the highly prevalent V599E mutant (in which the amino acid

valine, V, at position 599 in the protein is changed to glutamic acid, E).

Surprisingly, however, three mutants showed impaired, not enhanced, catalytic activity *in vitro* — but still stimulated ERK *in vivo*. How could these mutants switch ERK on? Remarkably, it seems that they do so by, in effect, playing tag, activating another RAF-family member, C-RAF (Fig. 1c). As has been reported previously¹¹, Wan and co-workers detected an association between B-RAF and C-RAF, which occurs whether B-RAF is normal or mutated. But C-RAF activity is increased only in cells that express the 'kinase-activated' or 'kinase-impaired' B-RAF mutants — and not in cells expressing normal B-RAF. Moreover, although both classes of mutant activate C-RAF, only the kinase-impaired proteins specifically require C-RAF in order to activate ERK (Fig. 1b, c), a finding confirmed in cell lines derived from human tumours with B-RAF mutations.

To gain further insight into the effects of these mutations, Wan *et al.* determined the structure of the normal and a cancer-associated B-RAF kinase domain — an impressive feat that researchers have struggled to accomplish for nearly a decade. As in other protein kinases, the B-RAF catalytic domain adopts a bi-lobal structure. Between the amino- and carboxy-terminal lobes is a cleft that binds the energy-providing molecule ATP.

Also consistent with other kinases, phosphorylation of the activation segment of the kinase domain is required for optimal catalysis — as noted above — and to stabilize the active shape of the protein. The crystal structure of the normal B-RAF kinase domain represents the inactive conformation. In this structure, the activation segment interacts with amino acids in a glycine-rich stretch — known as the P-loop — in the amino-terminal lobe. Strikingly, 70% of cancer-associated B-RAF mutations occur in either the P-loop or the activation segment, apparently destabilizing this inactive conformation. Some mutations, such as V599E, may inflict a double whammy by not only destabilizing the inactive structure but also mimicking the effect of phosphorylating the activation segment.

Interestingly, the mutations that impair kinase activity also localize to these regions, but instead they alter amino acids that are important in catalysis. Wan *et al.* propose that, like the kinase-activating mutations, these kinase-impairing changes destabilize the inactive structure, producing the active conformation. Crucially, however, the enzyme is still kinase-impaired, because amino acids involved in catalysis are mutated. Destabilization of the inactive structure could account for the cancer-promoting properties of this mutant class, but how it allows the impaired enzyme to activate the related C-RAF is unclear. Perhaps, by adopting an active structure, these mutants induce activating conformational changes in associated C-RAF proteins. Alternatively, they might recruit molecules that activate C-RAF, or titrate out negative regulators that inhibit it.

In this era of targeted therapeutics, these studies⁶ should provide crucial structural information to guide the development of more effective inhibitors of this enzyme family. Given that the promotion of tumour development by RAF proteins is a family affair — involving both B-RAF and C-RAF, which share highly related catalytic domains and downstream substrates — inhibitors that effectively block the activity of both proteins could hold great promise as broad-spectrum treatments for cancer.

Deborah K. Morrison is in the Laboratory of Protein Dynamics and Signaling, Center for Cancer Research, National Cancer Institute, Frederick, Maryland 21702-1201, USA.
e-mail: dmorrison@ncicrf.gov

1. Davies, H. *et al.* *Nature* **417**, 949–954 (2002).
2. Rajagopalan, H. *et al.* *Nature* **418**, 934 (2002).
3. Singer, G. *et al.* *J. Natl Cancer Inst.* **95**, 484–486 (2003).
4. Brose, M. S. *et al.* *Cancer Res.* **62**, 6997–7000 (2002).
5. Cohen, Y. *et al.* *J. Natl Cancer Inst.* **95**, 625–627 (2003).
6. Wan, P. T. C. *et al.* *Cell* **116**, 855–867 (2004).
7. Kolch, W. *Biochem. J.* **351**, 289–305 (2000).
8. Chong, H., Vikis, H. G. & Guan, K. L. *Cell Signal.* **15**, 463–469 (2003).
9. Mason, C. S. *et al.* *EMBO J.* **18**, 2137–2148 (1999).
10. Chong, H., Lee, J. & Guan, K. L. *EMBO J.* **20**, 3716–3727 (2001).
11. Weber, C. K. *et al.* *Cancer Res.* **61**, 3595–3598 (2001).

Obituary

John A. Pople (1925–2004)

Sir John Pople died on 15 March 2004 at the age of 78. He was a giant in his chosen field, computational quantum chemistry, for which he was awarded the Nobel Prize in Chemistry in 1998. Pople revolutionized the way chemistry is practised today by making it possible to carry out research in the computer as a complement to the conventional chemistry of the laboratory. It was Pople who was most responsible for making computational quantum chemistry available to the community of chemists at large, and he dominated the scene in this area during the past five decades.

Pople was born in Burnham-on-Sea in south-west England in 1925. His talents in mathematics were apparent at an early age. However, he describes in his Nobel autobiography how he introduced deliberate errors into his mathematics exercises at Bristol Grammar School so as not to appear too clever. It was not until a new mathematics teacher arrived, and set a particularly challenging test, that he succumbed to temptation and turned in a perfect paper, including multiple solutions to several of the problems. Despite the remarkable achievements that were to follow, he never lost his modest manner.

Pople went on to study mathematics and physics at the University of Cambridge. But eventually his interest in pure mathematics began to wane and he decided to apply his mathematical skills to chemistry. The germs of the ideas that were to provide the focus for his life's work — developing mathematical models that could describe all of chemistry — were sown in 1952. There were some digressions, including early seminal contributions and a classic text in the then-emerging field of nuclear magnetic resonance, work that was a forerunner to the magnetic resonance imaging of today. But the main game was always the creation of models for studying chemistry.

It was known long ago, and enunciated by Paul Dirac in 1929, that the laws of quantum mechanics could in principle be used to predict all of chemistry. What Pople did was convert this to a practical reality. His aim was to enable chemists to straightforwardly predict the properties of molecules, such as molecular structures and the way molecules react with one another, by using a computer rather than by carrying out experiments.

He went about this task by first formulating the essential characteristics of an acceptable theoretical model. For



Innovator in computational chemistry

example, it should be unique, well defined, unbiased, objective and widely applicable, as well as satisfying certain, more technical, requirements (such as being 'size-consistent'). If the model performed satisfactorily in systematic comparisons with experimental data, it could then be used to make predictions in cases where experimental data are not available.

Pople then proceeded to design a series of theoretical procedures that could serve as the basis for the model. In some cases, he was the inventor or co-inventor of the new procedures — for example, the Pariser–Parr–Pople π -electron theory of the 1950s, or 'complete neglect of differential overlap' (CNDO) and other 'semi-empirical' theories of the 1960s. In others, he reformulated existing ideas, such as Møller–Plesset perturbation theory, to make them practically and objectively applicable to a wide cross-section of chemistry.

A key to implementing such procedures was the construction of efficient computer programs, and this was one of Pople's major contributions. His development and release of the Gaussian 70 program marked a turning point in the field, making it possible to carry out broad theoretical studies on real chemical problems on a scale that had not previously been possible. This was also the hallmark of Pople's subsequent computational initiatives. Even in the more recently popularized density functional theory, where he was formally

only a minor player (the dominant contributor being Walter Kohn, with whom Pople shared the Nobel prize), he systematized the way such calculations were carried out and incorporated them into the Gaussian program. This undoubtedly helped significantly in accelerating the widespread acceptance and use of density functional theory in chemistry.

Computational quantum chemistry is used today by myriad chemists across a variety of fields. It has become a viable adjunct to experiment, and is used in solving fundamental problems. It is also being increasingly applied by industrial companies in more practical situations, such as the design of new drugs and new materials. It is especially valuable in studying substances that might be difficult to examine experimentally, for instance because they have a very short lifetime or are toxic or explosive. The computer is oblivious to such hazards.

The success of computational chemistry has of course been helped by massive and continuing increases in computer power. The computer available to Pople in 1970, when Gaussian 70 was born, was a CDC 1604. A standard personal computer today is roughly 100,000 times faster, has 5,000 times as much memory, and yet costs several hundred times less.

Apart from being a brilliant researcher, Pople was also a great communicator, as anyone who heard him lecture will know. Some people have the knack of making simple things complicated. Pople kept simple things simple, and had the gift of making complicated material appear simple as well.

Virtually all of Pople's prizewinning work was carried out while he was a professor at Carnegie–Mellon University in Pittsburgh, Pennsylvania, from 1964 to 1993, although it continued after he moved to Northwestern University in Evanston, Illinois. He also had close connections with the Australian National University in Canberra, which he described as his second academic home, and which he visited on nine occasions during the 1980s and 1990s. His death is a great loss not only for science but also for his many friends and colleagues.

John Pople was married to Joy Bowers in 1952. They shared a close relationship for just under 50 years before Joy passed away in 2002. He is survived by his four children, 11 grandchildren and one great-granddaughter.

Leo Radom

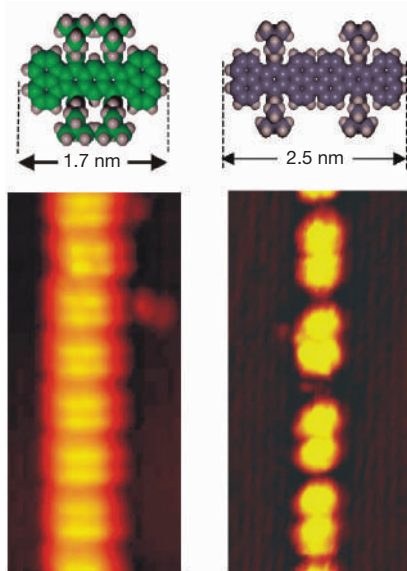
Leo Radom is in the School of Chemistry, University of Sydney, Sydney, New South Wales 2006, Australia.
e-mail: radom@chem.usyd.edu.au

Surface chemistry

The molecules have landed

Angew. Chem. Int. Edn **43**, 2092–2095 (2004)

The 'Lander' molecules got their name from their resemblance to the Mars Lander craft. They have a conducting 'board' and four insulating legs, which act as tiny pylons that support the conducting board and isolate it electrically from the surface below. Roberto Otero and colleagues have found that, on a copper oxide surface patterned with 2-nanometre-wide lines of copper, the length



of a Lander molecule determines whether it will orient across or along the copper lines.

Otero *et al.* report that short 'single' Landers ($C_{90}H_{98}$, above left) arrange themselves perpendicularly to the copper stripes, but longer 'violet' Landers ($C_{108}H_{104}$, above right) find it energetically more favourable to align along the copper lines. Such spatial and orientational control of these (and probably other) molecules could prove a convenient way to fabricate integrated electronic circuits on the nanoscale. It would allow molecules to be arranged in thermally stable chains without the need for direct manipulations with, for example, a scanning tunnelling microscope.

Rosamund Daw

Neurobiology

Mutation skews eyes

J. Neurophysiol. **91**, 2066–2078 (2004)

In people, defects in the cerebellum — a brain region involved in balance and coordination — commonly cause abnormal eye movements. In mice, mutations that alter a voltage-sensitive calcium channel, the P/Q channel, are predicted to interfere with signalling within and between cerebellar

neurons. So might these mutations also interfere with eye movements? John S. Stahl has found that they do, suggesting that the mutant mice may prove useful for studying the role of the cerebellum in controlling eye movement.

Using video recordings, Stahl measured reflex eye movements in so-called *rocker* mice, in which P/Q channels are mutated. He found substantial abnormalities compared with normal mice, such as lower amplitudes of movement and unusual vertical eye positions. Both humans and mice with P/Q mutations suffer age-related cerebellar degeneration, which could itself alter eye movements. But Stahl showed that some of the movement abnormalities are lifelong, and so could result directly from the effects of the inherited mutation on the electrical properties of the P/Q channel.

Laura Nelson

Astronomy

Star survey complete

Astron. Astrophys. doi:10.1051/0004-6361:20035959 (2004)

Basic information about our solar neighbourhood is still surprisingly incomplete, but the Geneva–Copenhagen survey of Sun-like stars in the Milky Way has now added an enormous amount of detail about our galaxy.

B. Nordström *et al.* have catalogued the precise three-dimensional motions, ages, temperatures and compositions of more than 14,000 stars. The research required roughly 63,000 individual velocity observations spread over 15 years, and includes almost all the Sun-like stars within 150 light years of Earth.

The survey reveals that about one third of the stars exist in binary systems, and that current models describing the dynamical heating of disk stars seem unable to match data that relate the ages and velocities of the stars. The team concludes that the history of our galaxy is much more turbulent than previously thought.

This mammoth data set is unlikely to be superseded before the results from the European Space Agency's GAIA satellite become available in 2015.

Mark Peplow

Cancer

A famine connection

J. Natl Cancer Inst. **96**, 539–546 (2004)

Animal studies have hinted that reducing food intake may lower the risk of breast cancer, although different dietary regimes have different effects. But Sjoerd G. Elias *et al.* suggest that short, severe bouts of food deprivation may in fact increase breast-cancer risk.

The authors studied a group of 15,000 Dutch women who were alive during the famine of 1944 to 1945 in the Netherlands.

Fifty-five years after the famine, 585 of these women were diagnosed with breast cancer. Elias *et al.* found that the risk of breast cancer correlated with the severity of famine exposure. Women who experienced the most severe food deprivation had a 48% higher risk of developing the condition than those unaffected by the famine. The association was even stronger for women who were aged between 2 and 9 when the food shortage hit.

The mechanisms that might link severe food deprivation to breast cancer are not known, but the authors suggest that several hormonal systems may be involved. These systems might adapt to famine and then be unable to deal with food abundance in later life.

Helen R. Pilcher

Metabolism

Is size everything?

Funct. Ecol. **18**, 184–187 (2004)

Bigger animals need to generate greater metabolic power than their smaller counterparts, but the relationship between size and metabolic performance is not linear. Rather, metabolic rate (Q) scales with body size (M) as $Q \propto M^b$. Researchers have long argued in favour of a universal value of b , but a fresh analysis suggests that there isn't one.

F. Bokma examined 3,572 individual measurements of metabolic rate and body size in fish, drawn from 217 separate data sets on 113 species, and encompassing temperatures from -1.5°C to 38°C . The average value of b was 0.715, but different groups showed their own conserved values. Most strikingly, a group of 37 measurements of sea trout (*Salmo trutta trutta*) yielded a statistically robust b of 0.86.

This is a far cry from previous pronouncements that b should hold steady throughout the animal kingdom. Theorists had suggested a constant b of 0.667, as would be expected if metabolism scales with body surface area, or of 0.75, as predicted by the supposedly fractal nature of internal transport systems. The latest analysis hints that current theories may well require an extra layer of complexity.

Michael Hopkin



M. LORENZONI

Deadly strike mechanism of a mantis shrimp

This shrimp packs a punch powerful enough to smash its prey's shell underwater.

Stomatopods (mantis shrimp) are well known for the feeding appendages they use to smash shells and impale fish. Here we show that the peacock mantis shrimp (*Odontodactylus scyllarus*) generates an extremely fast strike that requires major energy storage and release, which we explain in terms of a saddle-shaped exoskeletal spring mechanism. High-speed images reveal the formation and collapse of vapour bubbles next to the prey due to swift movement of the appendage towards it, indicating that *O. scyllarus* may use destructive cavitation forces to damage its prey.

Stomatopod appendages were previously thought to be limited to a maximum speed of 10 m s^{-1} (ref. 1) but, by using new imaging technology and a faster species, we have measured the dactyl heel reaching peak speeds of $14\text{--}23 \text{ m s}^{-1}$, peak angular speeds of $670\text{--}990 \text{ rad s}^{-1}$, and peak acceleration of $65\text{--}104 \text{ km s}^{-2}$ within an average period of 2.7 ms, making *O. scyllarus* perhaps the fastest appendicular striker in the animal kingdom.

To generate these extreme movements in water, a large amount of energy must be released over a short period. Stomatopods increase their power output through the use of a click mechanism, in which latches prevent appendage movement until muscle contraction is maximal^{1,2}. When the latch is freed, stored energy is released over a shorter time than the duration of the original muscle contraction. Some extreme animal movements also require a specialized spring, because the muscle fibres and tendon store insufficient elastic strain energy^{3–5}. We conservatively calculate that a minimum power requirement of 4.7×10^5 watts per kilogram of muscle is necessary for a typical strike, which is orders of magnitude higher than that available in the fastest-contracting muscles^{3,6} known. Moreover, the lateral extensor muscle and apodeme (arthropod tendon) could store only a small

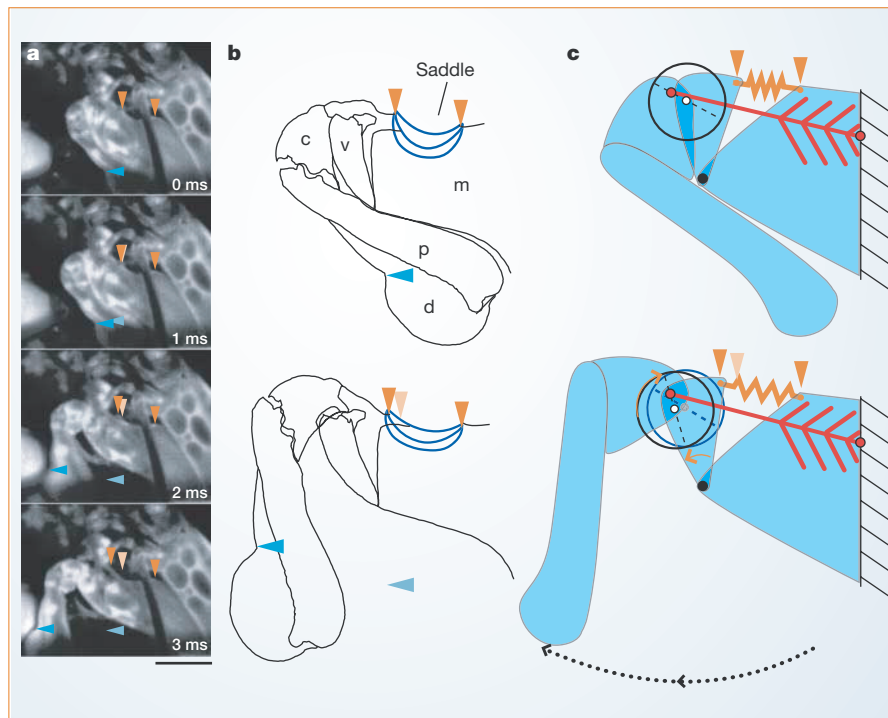


Figure 1 The mechanics of a stomatopod strike. **a**, A high-speed image sequence illustrates the distal extension of the saddle (orange triangles) occurring simultaneously with the extension of the smashing heel of the dactyl and propodus (blue triangles). Scale bar, 1 cm. Appendage kinematics were based on 6 individuals and 7–12 strikes per individual; measurement error, $\pm 4\%$. Images were recorded at 5,000 frames s^{-1} . **b**, The compressed (top) and released (bottom) saddle on the raptorial appendage (raptorial segments: m, merus; v, meral-V; c, carpus; p, propodus; d, dactyl). **c**, The saddle modelled as a spring (orange) that stores elastic energy to drive the movement. Top, pre-strike phase. The lateral extensor muscle and apodeme (red) pull on the carpus to compress the saddle, while flexor muscles (not shown) engage a click mechanism to prevent extension of the appendage¹. Bottom, the strike occurs when the latch is released¹, enabling the saddle to extend and two pivot points to rotate in opposite directions. The meral-V forms a pivot point (black circle) as it rotates distal-ventrally (anticlockwise here) and pushes the second pivot point, the carpal-meral fulcrum (white circle), distally. The isometrically contracted extensor muscle maintains a constant distance between its carpal and meral attachment points (red circles), forcing the carpus to rotate (clockwise here) and driving the dactyl heel towards the prey.

fraction of the strain energy underlying the work requirements in an average strike (see supplementary information), which means that stomatopods need a specialized spring.

In our model for the stomatopod strike, elastic energy is stored in a compressive, saddle-shaped spring (Fig. 1) that is a stiff exoskeletal structure located dorsally on the merus (the enlarged proximal segment) of all stomatopods. Hyperbolic-paraboloid (saddle-shaped or anticlastic) surfaces are used in engineering and architecture: their opposite and transverse curvatures reduce failure by distributing stresses across the three-dimensional surface. Likewise, the saddle shape of the stomatopod's spring minimizes the probability of local buckling while compressing and extending. To our knowledge, the stomatopod's saddle is the first biological hyperbolic-paraboloid spring to be described.

Stomatopod smashing produces a loud pop, and high-speed video images reveal evidence of cavitation (Fig. 2). Cavitation

occurs in fluids when areas of low pressure form vapour bubbles that collapse and yield considerable energy (in the form of heat, light and sound); these can be sufficient to destroy boat propellers and other hard surfaces⁷. Cavitation has been noted in snapping shrimp, which appear to 'shoot' cavitation bubbles at prey items to stun them^{8,9}.

By contrast, cavitation in stomatopods occurs between the surface being struck and the dactyl heel. Although the heel is highly mineralized¹⁰, the surface becomes pitted and damaged over time; stomatopods moult frequently and produce a new smashing surface every few months. In addition to their novel energy-storage mechanism and remarkable striking speed, stomatopods may be using cavitation forces to process their prey.

S. N. Patek, W. L. Korff, R. L. Caldwell

Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA
e-mail: patek@socrates.berkeley.edu

1. Burrows, M. *Zeit. Vergl. Physiol.* **62**, 361–381 (1969).

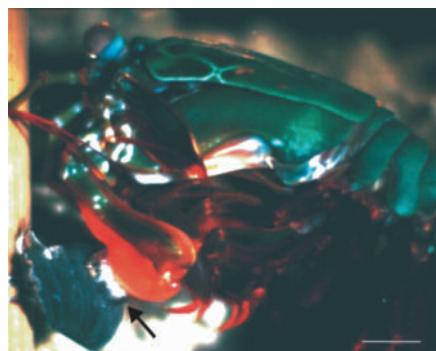


Figure 2 The mantis shrimp *Odontodactylus scyllarus* strikes a snail. Cavitation bubbles (arrow) form and collapse between the dactyl heel (red) and the snail (black). Scale bar, 1 cm.

2. Burrows, M. & Hoyle, G. *J. Exp. Zool.* **179**, 379–394 (1972).
3. Alexander, R. M. & Bennet-Clark, H. C. *Nature* **265**, 114–117 (1977).
4. Bennet-Clark, H. C. in *The Insect Integument* (ed. Hepburn, H. R.) 421–443 (Elsevier, Amsterdam, 1976).
5. Gronenberg, W. J. *Comp. Phys. A* **178**, 727–734 (1996).
6. Alexander, R. M. *Comp. Biochem. Phys. A* **133**, 1001–1011 (2002).
7. Brennen, C. E. *Cavitation and Bubble Dynamics* (Oxford University Press, New York, 1995).
8. Lohse, D., Schmitz, B. & Versluis, M. *Nature* **413**, 477–478 (2001).
9. Versluis, M., Schmitz, B., von der Heydt, A. & Lohse, D. *Science* **289**, 2114–2117 (2000).
10. Currey, J. D., Nash, A. & Bonfield, W. J. *Mater. Sci.* **17**, 1939–1944 (1982).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Origin of AIDS

Contaminated polio vaccine theory refuted

Despite strong evidence to the contrary^{1–5}, speculation continues that the AIDS virus, human immunodeficiency virus type 1 (HIV-1), may have crossed into humans as a result of contamination of the oral polio vaccine (OPV)^{6–8}. This 'OPV/AIDS theory' claims that chimpanzees from the vicinity of Stanleyville — now Kisangani in the Democratic Republic of Congo — were the source of a simian immunodeficiency virus (SIVcpz) that was transmitted to humans when chimpanzee tissues were allegedly used in the preparation of OPV^{6,7}. Here we show that SIVcpz is indeed endemic in wild chimpanzees of this region but that the circulating virus is phylogenetically distinct from all strains of HIV-1, providing direct evidence that these chimpanzees were not the source of the human AIDS pandemic.

Detection and molecular characterization of SIVcpz in chimpanzee communities in the vicinity of Kisangani should directly test the OPV/AIDS theory. An earlier survey of chimpanzees at Wanie-Rukula near Kisangani (Fig. 1a; W. D. Hamilton, M. W. and J. B. J., January 2000, see ref. 9) failed to identify SIVcpz viral (v) RNA in any of 34 faecal samples collected. However, western immunoblot analysis of 10 chimpanzee urine samples collected at the same time identified two specimens that showed strong crossreactivity with the HIV-1 core protein p24 (Fig. 1b). Such indeterminate urine antibody profiles were found in chimpanzees from Tanzania, where SIVcpz infection was subsequently demonstrated after amplification by polymerase chain reaction (PCR) and sequencing of DNA¹⁰.

To confirm the existence of SIVcpz in the Kisangani apes and to identify circulating strain(s) at a molecular level, we resumed field-work in February 2003, this time collecting 97 faecal samples from three different sites (for map, see supplementary information). From these, we identified one SIVcpz vRNA-

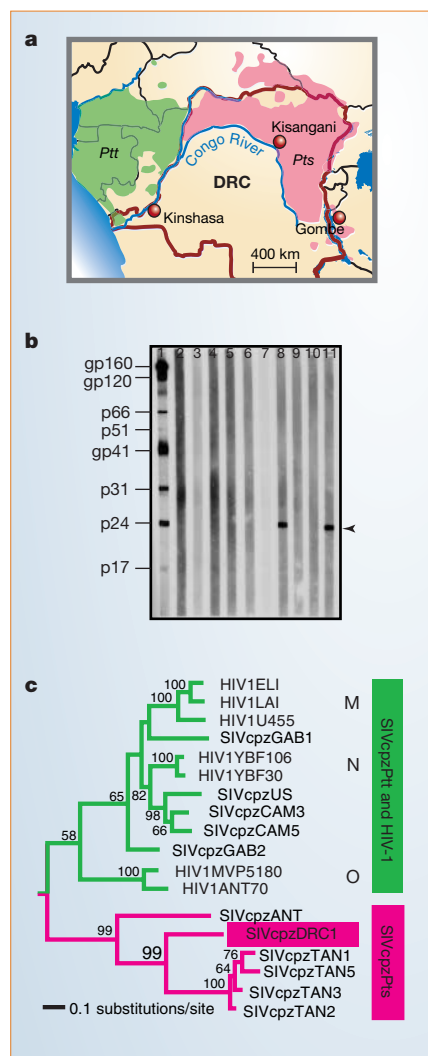


Figure 1 The Kisangani expeditions and refutation of the OPV/AIDS theory. **a**, Map of the Democratic Republic of Congo (DRC) and neighbouring countries showing the ranges of *Pan troglodytes troglodytes* (Ptt) and *Pan troglodytes schweinfurthii* (Pts). **b**, Western blot analysis of urine samples from 10 chimpanzees, collected during the 2000 expedition to rainforests near Kisangani; plasma from a positive HIV-1 control is shown in the left lane, with cross-reactivity to the different viral proteins indicated. Two samples showed strong cross-reactivity with HIV-1 p24 (arrowhead). **c**, Maximum-likelihood-estimated phylogenetic tree for gp41/nef sequence data, with bootstrap results. Bootstrap percentages are shown for all clades that were present in more than 50% of 1,000 maximum-likelihood-inferred bootstrap trees. Support for the SIVcpzDRC1/SIVcpzTAN clade was very strong (99%). Analysis of partial gag sequences supported the same phylogenetic position for SIVcpzDRC1 (data not shown). For further details of phylogenetic analyses, see supplementary information.

positive specimen from the Parisi forest by PCR amplification of gag (422 base pairs) and gp41/nef (699 base pairs) sequences. This result confirmed that natural SIVcpz infection was present in chimpanzees in the Kisangani region.

Phylogenetic analysis of the newly derived sequences revealed that the Kisangani virus clustered with high statistical support with SIVcpz strains that were infecting chimpanzees of the same subspecies (*Pan*

troglodytes schweinfurthii) that lived about 800 km to the south-east in Gombe National Park in Tanzania^{10,11}. The new virus, which we designate SIVcpzDRC1, represents a third lineage within the well circumscribed *P. t. schweinfurthii* SIVcpz radiation, and is clearly distinct from the *P. t. troglodytes* SIVcpz clade that includes all known strains of HIV-1 (Fig. 1c, and see supplementary information).

These results indicate that chimpanzees in the vicinity of Kisangani are endemically infected with SIVcpz that is highly divergent from HIV-1, thereby ruling out these apes as the source of HIV-1 and refuting the OPV/AIDS theory. Instead, each of the many circulating HIV-1 variants comprising groups M, N and O is linked to SIVcpz from *P. t. troglodytes* (Fig. 1c), the chimpanzee subspecies native to west-central Africa^{1,12}.

Given that fears about the safety of polio vaccines are currently threatening the global campaign to eradicate the disease⁸, our clear-cut evidence against one of the key sources of concern is timely. The molecular epidemiological data presented here, together with data suggesting that HIV-1 group M originated 30 years before OPV trials were conducted^{1,13} and the absence of detectable SIVcpz or chimpanzee DNA in archival stocks of OPV^{2,3}, should finally lay the OPV/AIDS theory to rest.

Michael Worobey*, **Mario L. Santiago†**, **Brandon F. Keele†**, **Jean-Bosco N. Ndjango‡**, **Jeffrey B. Joy§**, **Bernard L. Labamall**, **Benoit D. Dheda†**, **Andrew Rambaut¶**, **Paul M. Sharp#**, **George M. Shaw†**, **Beatrice H. Hahn†**

*Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, USA e-mail: worobey@email.arizona.edu

†Departments of Medicine and Microbiology, and ‡Howard Hughes Medical Institute, University of Alabama, Birmingham, Alabama 35294, USA

§Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

¶Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

#Institute of Genetics, University of Nottingham, Nottingham NG7 2UH, UK

1. Sharp, P. M. *et al.* *Phil. Trans. R. Soc. Lond. B* **356**, 867–876 (2001).
2. Blancou, P. *et al.* *Nature* **410**, 1045–1046 (2001).
3. Berry, N. *et al.* *Nature* **410**, 1046–1047 (2001).
4. Rambaut, A., Robertson, D. L., Pybus, O. G., Peeters, M. & Holmes, E. C. *Nature* **410**, 1047–1048 (2001).
5. Plotkin, S. A. *Phil. Trans. R. Soc. Lond. B* **356**, 815–823 (2001).
6. Hooper, E. *The River: A Journey Back to the Source of HIV and AIDS* (Penguin, London, 1999).
7. Hooper, E. *Atti Conv. Lincei* **187**, 27–230 (2003).
8. Butler, D. *Nature* **428**, 109 (2004).
9. Trivers, R. *Nature* **404**, 828 (2000).
10. Santiago, M. L. *et al.* *J. Virol.* **77**, 2233–2242 (2002).
11. Santiago, M. L. *et al.* *Science* **295**, 465 (2001).
12. Gao, F. *et al.* *Nature* **397**, 436–441 (1999).
13. Korber, B. *et al.* *Science* **288**, 1789–1796 (2000).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

The worldwide leaf economics spectrum

Ian J. Wright¹, Peter B. Reich², Mark Westoby¹, David D. Ackerly³, Zdravko Baruch⁴, Frans Bongers⁵, Jeannine Cavender-Bares⁶, Terry Chapin⁷, Johannes H. C. Cornelissen⁸, Matthias Diemer⁹, Jaime Flexas¹⁰, Eric Garnier¹¹, Philip K. Groom¹², Javier Gulias¹⁰, Kouki Hikosaka¹³, Byron B. Lamont¹², Tali Lee¹⁴, William Lee¹⁵, Christopher Lusk¹⁶, Jeremy J. Midgley¹⁷, Marie-Laure Navas¹¹, Ülo Niinemets¹⁸, Jacek Oleksyn^{2,19}, Noriyuki Osada²⁰, Hendrik Poorter²¹, Pieter Poot²², Lynda Prior²³, Vladimir I. Pyankov²⁴, Catherine Roumet¹¹, Sean C. Thomas²⁵, Mark G. Tjoelker²⁶, Erik J. Veneklaas²² & Rafael Villar²⁷

¹Department of Biological Sciences, Macquarie University, New South Wales 2109, Australia

²Department of Forest Resources, University of Minnesota, St Paul, Minnesota 55108, USA

³Department of Biological Sciences, Stanford University, Stanford, California 94305, USA

⁴Departamento de Estudios Ambientales, Universidad Simón Bolívar, Caracas 1080, Venezuela

⁵Forest Ecology and Forest Management Group, Department of Environmental Sciences, Wageningen University, PO Box 342, 6700 AH Wageningen, The Netherlands

⁶Smithsonian Environmental Research Center, PO Box 28, 647 Contees Wharf Road, Edgewater, Maryland 21037, USA

⁷Institute of Arctic Biology, University of Alaska, Fairbanks, Alaska 99775, USA

⁸Institute of Ecological Science, Department of Systems Ecology, Vrije Universiteit, De Boelelaan 1087, 1081 HV, Amsterdam, The Netherlands

⁹Institute für Umweltwissenschaft, University of Zurich, Zurich, Switzerland

¹⁰Departament de Biologia, Laboratori de Fisiologia Vegetal, Universidad de Illes Balears, 07122 Palma de Mallorca, Illes Balears (Spain)

¹¹Centre d'Ecologie Fonctionnelle et Evolutive, CNRS, UMR 5175, 1919, Route de Mende, 34293 Montpellier cedex 5, France

¹²Department of Environmental Biology, Curtin University of Technology, Perth, Western Australia 6845, Australia

¹³Graduate School of Life Sciences, Tohoku University, Aoba, Sendai 980-8578, Japan

¹⁴Department of Biology, University of Wisconsin-Eau Claire, Eau Claire, Wisconsin 54702-4004, USA

¹⁵Landcare Research, Private Bag 1930, Dunedin, New Zealand

¹⁶Departamento de Botánica, Universidad de Concepción, Casilla 160-C, Concepción, Chile

¹⁷Department of Botany, University of Cape Town, ZA-7701 Rondebosch, South Africa

¹⁸Department of Plant Physiology, University of Tartu, Riia 23, Tartu 51011, Estonia

¹⁹Polish Academy of Sciences, Institute of Dendrology, Parkowa 5, 62-035 Kornik, Poland

²⁰Nikko Botanical Garden, Graduate School of Science, University of Tokyo, 1842 Hanaishi, Nikko, Tochigi 321-1435, Japan

²¹Plant Ecophysiology, Utrecht University, PO Box 800.84, 3508 TB, Utrecht, The Netherlands

²²School of Plant Biology, University of Western Australia, Crawley, Western Australia 6009, Australia

²³Key Centre for Tropical Wildlife Management, Charles Darwin University, Darwin, Northern Territory 0909, Australia

²⁴Ural State University, Yekaterinburg, Russia

²⁵Faculty of Forestry, University of Toronto, 33 Willcocks St, Toronto, Ontario M5S 3B3, Canada

²⁶Department of Forest Science, Texas A&M University, 2135 TAMU, College Station, Texas 77843-2135, USA

²⁷Area de Ecología, Campus de Rabanales, Universidad de Córdoba, 14071 Córdoba, Spain

Bringing together leaf trait data spanning 2,548 species and 175 sites we describe, for the first time at global scale, a universal spectrum of leaf economics consisting of key chemical, structural and physiological properties. The spectrum runs from quick to slow return on investments of nutrients and dry mass in leaves, and operates largely independently of growth form, plant functional type or biome. Categories along the spectrum would, in general, describe leaf economic variation at the global scale better than plant functional types, because functional types overlap substantially in their leaf traits. Overall, modulation of leaf traits and trait relationships by climate is surprisingly modest, although some striking and significant patterns can be seen. Reliable quantification of the leaf economics spectrum and its interaction with climate will prove valuable for modelling nutrient fluxes and vegetation boundaries under changing land-use and climate.

Green leaves are fundamental for the functioning of terrestrial ecosystems. Their pigments are the predominant signal seen from space. Nitrogen uptake and carbon assimilation by plants and the decomposability of leaves drive biogeochemical cycles. Animals, fungi and other heterotrophs in ecosystems are fuelled by photosynthate, and their habitats are structured by the stems on which leaves are deployed. Plants invest photosynthate and mineral nutrients in the construction of leaves, which in turn return a revenue stream of photosynthate over their lifetimes. The photosynthate is used to acquire mineral nutrients, to support metabolism and to re-invest in leaves, their supporting stems and other plant parts.

There are more than 250,000 vascular plant species, all engaging in the same processes of investment and reinvestment of carbon and mineral nutrients, and all making enough surplus to ensure continuity to future generations. These processes of investment and re-investment are inherently economic in nature^{1–3}. Understanding how these processes vary between species, plant functional types

and the vegetation of different biomes is a major goal for plant ecology and crucial for modelling how nutrient fluxes and vegetation boundaries will shift with land-use and climate change.

Data set and parameters

We formed a global plant trait network (Glopanet) to quantify leaf economics across the world's plant species. The Glopanet data set spans 2,548 species from 219 families at 175 sites (approximately 1% of the extant vascular plant species). The coverage of traits, species and sites is at least tenfold greater than previous data compilations^{4–11}, extends to all vegetated continents, and represents a wide range of vegetation types, from arctic tundra to tropical rainforest, from hot to cold deserts, from boreal forest to grasslands. Site elevation ranges from below sea level (Death Valley, USA) to 4,800 m. Mean annual temperature (MAT) ranges from -16.5°C to 27.5°C ; mean annual rainfall (MAR) ranges from 133 to 5,300 mm per year. This covers most of the range of MAT–MAR space in which higher plants occur¹² (Fig. 1). The broad coverage of the data set has

allowed us to quantify the relationships of leaf economics to climate at a scale not previously possible. Here we report some global outcomes from our analyses.

We focus on six key features of leaves that together capture many essentials of leaf economics. (1) Leaf mass per area (LMA) measures the leaf dry-mass investment per unit of light-intercepting leaf area deployed. Species with high LMA have a thicker leaf blade or denser tissue, or both. (2) Photosynthetic assimilation rates measured under high light, ample soil moisture and ambient CO_2 are here called photosynthetic capacity (A_{mass}) for brevity. Photosynthetic capacity is influenced both by stomatal conductance and by the drawdown of CO_2 concentration inside the leaf (carboxylation capacity). (3) Leaf nitrogen (N) is integral to the proteins of photosynthetic machinery, especially Rubisco^{8,13}. The photosynthetic machinery is responsible for drawdown of CO_2 inside the leaf, a process also affected by leaf structure^{14,15}. (4) Leaf phosphorus (P) is found in nucleic acids, lipid membranes and bioenergetic molecules such as ATP. Phosphorus derives from weathering of soil minerals at a site, in contrast to nitrogen, much of which may be fixed from the atmosphere by plants. (5) Dark respiration rate (R_{mass}) reflects metabolic expenditure of photosynthate in the leaf, especially protein turnover and phloem-loading of photosynthates¹⁶. (6) Leaf lifespan (LL) describes the average duration of the revenue stream from each leaf constructed. Long LL requires

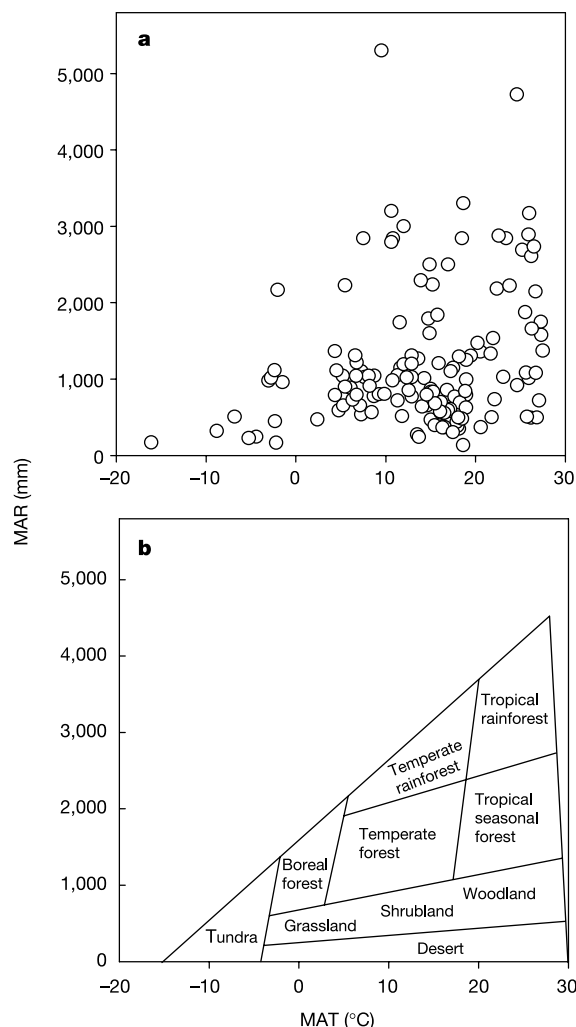


Figure 1 Mean annual rainfall (MAR) and mean annual temperature (MAT). Results for the 175 sites from where leaf data were compiled (**a**), in relation to major biome types of the world (**b**), following ref. 12. Biome boundaries are only approximate.

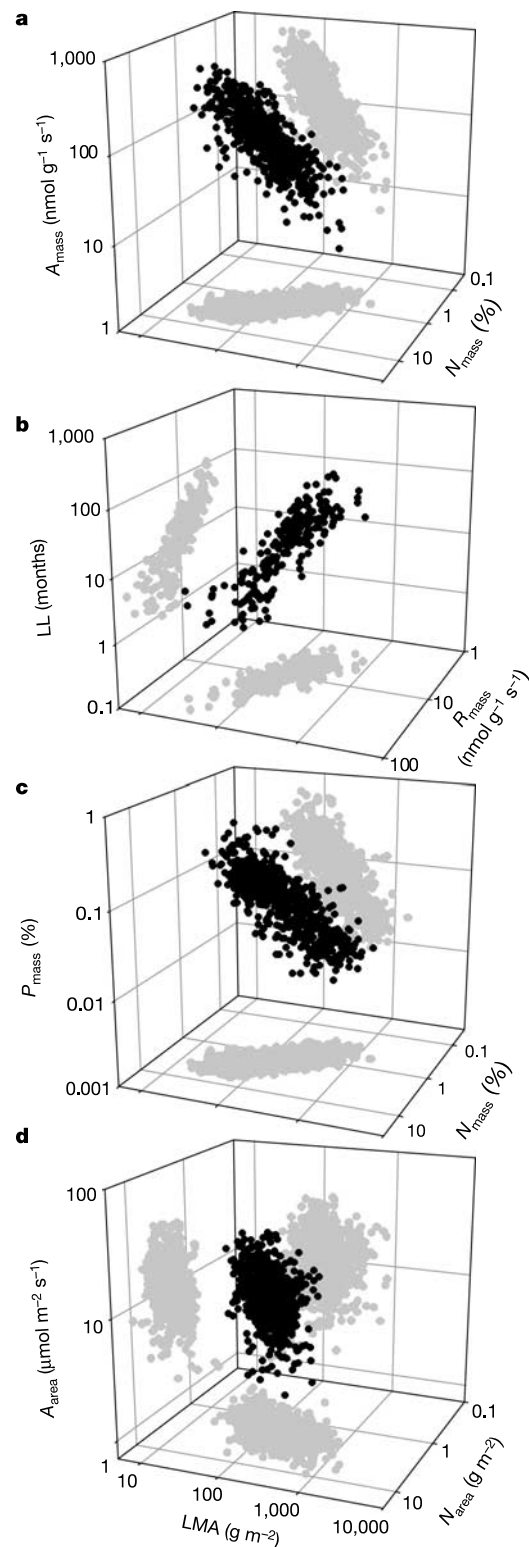


Figure 2 Three-way trait relationships among the six leaf traits with reference to LMA, one of the key traits in the leaf economics spectrum. The direction of the data cloud in three-dimensional space can be ascertained from the shadows projected on the floor and walls of the three-dimensional space. Sample sizes for three-way relationships are necessarily a subset of those for each of the bivariate relationships. **a**, A_{mass} , LMA and N_{mass} ; 706 species. **b**, LL, R_{mass} and LMA; 217 species. **c**, N_{mass} , P_{mass} and LMA; 733 species. **d**, A_{area} , LMA and N_{area} ; 706 species.

robust construction in the form of high LMA^{6,17,18}.

Ecophysiological attributes such as leaf N, leaf P, A_{mass} or R_{mass} can be expressed per leaf area, per leaf dry mass or per leaf volume. A leaf-area basis reflects fluxes in relation to surfaces. It is a natural basis for describing light capture and for expressing transactions through surfaces such as trade-offs between carbon gain and water transpiration^{19,20}. On a mass basis, leaf economics are quantified in terms of revenues and expenditures per unit investment, measured as biomass, or C, N or P. Scaling up to whole plants, mass-based expressions of leaf nutrient concentrations (N_{mass} , P_{mass}), A_{mass} and R_{mass} are more tightly correlated than area-based expressions to relative growth rates of seedlings or to absolute height growth rates of young trees^{21,22}. By definition, area- and mass-based traits can be interconverted via LMA (for example, $N_{\text{area}} = N_{\text{mass}} \times \text{LMA}$). Here we report relationships among leaf traits using both area- and mass-based formulations, and also analyses where the effect of LMA is explicitly controlled.

We first quantify relationships among the six mass-based leaf traits: A_{mass} , R_{mass} , LMA, LL, N_{mass} and P_{mass} . We find that these economic traits covary tightly across all species. Trait relationships are similar for species pooled by growth form, plant functional group or biome, which indicates the existence of a single global spectrum of leaf economic variation. Second, we treat photosynthetic capacity, R_{mass} and leaf nutrient concentrations on a leaf-area basis. These relationships are not as strong as among mass-basis traits, and we consider why. Third, we assess the influence of climate on leaf trait relationships. We find that, in general, the influence is modest, although particular traits and trait relationships show substantial patterning with climate.

The leaf economics spectrum

Mass-based leaf traits

The six mass-basis leaf traits varied by one to two orders of magnitude across the data set. LMA ranged from 14 to 1,500 g m⁻² and LL from 0.9 up to 288 months. A_{mass} ranged from 5 to 660 nmol g⁻¹ s⁻¹; dark respiration from 2.2 to 65 nmol g⁻¹ s⁻¹. N_{mass} ranged from 0.2 to 6.4%; P_{mass} from 0.008 to 0.6%. Considered pairwise, all leaf traits were highly correlated (Table 1). These correlations have been reported previously from smaller data sets^{6–10,23}. Here we have generalized the patterns over many more species, sites and vegetation types.

We moved beyond pairwise consideration of traits to determine the extent to which leaf economic traits covary in multidimensional trait space. This covariation can be quantified as the proportion of total trait variation explained by the first principal axis in a principal components analysis. In two-trait space, the principal axis is the long axis of the ellipse resulting from two correlated traits. In three-trait space, the principal axis is the long axis of an ellipsoid. In multi-dimensional trait space, the principal axis describes the main axis of variation through a hyperellipsoid²⁴. A remarkable 82% of all variation in A_{mass} , LMA and N_{mass} across species lay along the first principal axis in three-trait space (Fig. 2a). Because some of the residual 18% must be measurement variation, 82% represents a minimum estimate of the dominance of this single spectrum in explaining variation across plant species worldwide. Further three-

dimensional subsets of the six-dimensional data set are shown in Figs 2b and c. Multi-dimensional analyses including from four to all six of the traits similarly showed the large majority of variation explained with a single axis (Table 2). With the six traits included, 74% of all variation lay along the first principal axis.

The extent to which each trait contributed to the principal axis of variation is indicated by a loading (or weight) assigned to each trait. The directionality of these loadings (Table 2) indicates that the axis can be thought of as a leaf economics spectrum. This spectrum runs from species with potential for quick returns on investments of nutrients and dry mass in leaves to species with a slower potential rate of return. At the quick-return end are species with high leaf nutrient concentrations, high rates of photosynthesis and respiration, short leaf lifetimes and low dry-mass investment per leaf area. At the slow-return end are species with long leaf lifetimes, expensive high-LMA leaf construction, low nutrient concentrations, and low rates of photosynthesis and respiration.

Within growth forms or functional groups the principal axes of variation had the same directionality of trait correlations as for the total data set (Table 2). Similarly, species grouped by major biome type (Fig. 1), or by MAT or MAR classes, yielded the same pattern (data not shown). The concordance of these results is of special significance, indicating a coordination of these key leaf traits that is consistent across major plant functional types, growth forms and biomes. The amount of variation captured by the principal axis in the different species groupings was also similar to that across all species in most cases.

The main exception was among deciduous trees and shrubs where, as expected, there was substantially less variation in LL (5-fold versus 100-fold), and where LL–LMA relationships were partially uncoupled. Still, whereas different growth forms and functional groups were differentiated along the leaf economics spectrum when trait means were considered, the overlap between species groups was large (data not shown). Evergreen trees and shrubs had longer mean LL and higher LMA than deciduous species, but evergreens had much wider ranges for both traits, extending to LLs almost as short as for the shortest-LL deciduous species, and to similarly low LMA. Similarly, on average, shrubs and trees had higher LMA and longer LLs but lower N_{mass} , A_{mass} and R_{mass} than herbs and grasses, yet trees and shrubs spanned almost the entire range of any of these leaf traits. Another example: N_2 -fixing species had higher mean N_{mass} than non N_2 -fixing plants, yet the range of N_{mass} was larger and extended higher in non N_2 -fixing species.

Allometries among traits

Slopes on log–log axes, or ‘scaling exponents’, indicate the proportionality of pairwise trait relationships. Most slopes were significantly different from +1.0 or –1.0 (Table 1, above diagonal). In other words, the traits showed allometric relationships rather than scaling in direct proportion with one another (‘isometry’). A tenfold increase in P_{mass} corresponded with a 4.7-fold increase in N_{mass} (scaling slope 0.67), indicating that N:P ratios decline as one moves towards the end of the spectrum that represents quick returns on investments of carbon and nutrients. The stoichiometry between

Table 1 Mass basis of bivariate relationships between the leaf traits

	log LL	log LMA	log A_{mass}	log N_{mass}	log P_{mass}	log R_{mass}
log LL		1.71 (1.62, 1.82)	–1.38 (–1.45, –1.31)	–2.26 (–2.39, –2.14)	–1.06 (–1.19, –0.94)	–1.67 (–1.82, –1.53)
log LMA	0.42 (678)		–0.75 (–0.79, –0.72)	–1.28 (–1.32, –1.24)	–0.82 (–0.86, –0.78)	–0.96 (–1.05, –0.88)
log A_{mass}	0.68 (512)	0.50 (764)		1.72 (1.63, 1.81)	1.03 (0.91, 1.16)	1.18 (1.09, 1.27)
log N_{mass}	0.42 (706)	0.57 (1958)	0.53 (712)		0.66 (0.64, 0.69)	0.70 (0.64, 0.76)
log P_{mass}	0.24 (207)	0.55 (739)	0.16 (212)	0.72 (745)		1.04 (0.87, 1.25)
log R_{mass}	0.60 (217)	0.45 (274)	0.59 (259)	0.55 (267)	0.34 (78)	

Standardized major axis slopes with 95% confidence intervals are given in the upper right section of the matrix (y variable is column 1, x variable in row 1). Coefficients of determination (r^2) and sample sizes are given in the lower left section of the matrix. All relationships were highly significant, $P < 0.0001$. Further details allowing calculation of predictive regression equations for each pair of leaf traits are given in Supplementary Information.

Table 2 Principal components analyses of global leaf trait data

	All species	Trees*	Shrubs	Herbs*	Grasses*	N ₂ -fixers*	Non N ₂ -fixers	C3	C4†	Broad-leaved shrubs and trees	Needle-leaved shrubs and trees*	Evergreen shrubs and trees	Deciduous shrubs and trees*
Variation explained (%)	74.4	77.2	72.0	67.8	70.7	79.6	75.0	73.0	68.7	72.4	58.0	64.0	48.5
Leaf trait	Loadings												
LL	−0.85	−0.90	−0.83	−0.73	−0.86	−0.86	−0.84	−0.88		−0.84	−0.79	−0.74	−0.43
LMA	−0.88	−0.84	−0.84	−0.72	−0.91	−0.87	−0.87	−0.84	−0.84	−0.84	−0.73	−0.82	−0.28
N _{mass}	0.91	0.82	0.91	0.89	0.71	0.88	0.92	0.90	0.71	0.90	0.71	0.87	0.81
A _{mass}	0.86	0.91	0.82	0.88	0.79	0.96	0.86	0.88	0.87	0.83	0.82	0.76	0.84
R _{mass}	0.88	0.93	0.83	0.89	0.93	0.89	0.89	0.90	0.88	0.84	0.74	0.82	0.90
P _{mass}	0.79		0.87				0.80	0.72		0.85		0.78	

The principal axis or component explained 74.4% of variation in the total data set. LL and LMA were negatively correlated with this primary axis of variation while the other traits were positively correlated with it, both in the total data set and in data subsets defined by growth form or functional group. The same directionality of trait loadings and similarly high percentage of variance was explained by the principal axis with species grouped by site temperature, rainfall or altitude, or with sites grouped into major biome type following Fig. 1 (data not shown), demonstrating the broad generality of the coordinated spectrum of leaf economics. All data were log₁₀-transformed before analyses.

*P_{mass} excluded due to too few data.

†P_{mass} and LL excluded due to few data. C3, C4 indicate species with C3 and C4 photosynthetic pathways, respectively.

leaf N and P has begun to attract increasing interest as an index of soil nutrient limitation (taken across sets of co-occurring species)²⁵, and also because it relates to plant growth strategies and influences plant–herbivore interactions in food webs²⁶.

The slope of the LL–LMA relationship was 1.7, meaning that tenfold greater dry mass invested per unit leaf area coincided with 50-fold longer LL. All else being equal, this implies that the measure of light-intercepting leaf area (in mm²) × duration (in months) per gram leaf was greater for high-LMA than for low-LMA species. If this measure translated directly into fitness benefit, this might lead to runaway selection towards ever-increasing LMA and LL²⁷. But continuing ecological success of low-LMA species shows that all else is not equal. On average, a tenfold decrease in LMA, for example, coincided with a 21-fold increase in photosynthetic capacity. Further, low LMA, high A_{mass} and generally faster turnover of plant parts permit a more flexible response to the spatial patchiness of light and soil resources²⁸, as well as conferring advantages via a compound interest effect, whereby carbon fixed earlier can be reinvested in new leaves sooner^{27,29}. On the other hand, high A_{mass} requires high N_{mass}, and the combination of high LMA and high N_{mass} may increase vulnerability to herbivory as well as increasing energy losses via respiration⁹, which can be detrimental in situations where energy gain is low owing to low resource availability, such as under low light conditions³⁰.

Area versus mass basis of expression

Because N_{mass} = N_{area}/LMA, we considered whether the relationship between N_{mass} and LMA should be thought of as arising from N_{area} values that do not vary greatly across species, divided through by highly varying LMA. The evidence contradicts this interpretation. First, N_{area} varied more widely than did N_{mass} (35-fold versus 26-fold; data for 1,958 species). Second, the −1.28 slope of the N_{mass}–LMA relationship was significantly steeper (Table 1) than the −1.0 slope expected if N_{area} was independent of LMA.

N_{area} was indeed correlated with LMA, but positively, and more weakly than were N_{mass} and LMA ($r^2 = 0.34$ versus 0.57; Table 3). Gradients of leaf N on a mass versus an area basis hence represent

fundamentally different multiple-trait gradients because of their different patterns of covariation with LMA³¹. The leaf economics spectrum of species from low to high N_{mass} also constitutes a spectrum of decreasing LMA and LL, and of increasing A_{mass}, R_{mass} and P_{mass}. But a spectrum of leaf types in terms of N_{area} would be less informative. A given N_{area} can result from low N_{mass} combined with high LMA, high N_{mass} combined with low LMA, or from combinations in between. In general, low N_{mass} with high LMA represents a species with long-lived leaves and low A_{mass}, while high N_{mass} with low LMA represents the opposite. As a result, a wider variety of leaf types may be found at a given N_{area} than at a given N_{mass}^{31–33}. Here for example, LMA varied approximately 20-fold at the grand mean of N_{area} versus tenfold at the grand mean of N_{mass}.

Because of the covariation between leaf N and LMA, relationships between leaf N and other traits changed substantially when expressed on an area rather than on a mass basis. As seen previously^{7–9,31,33}, relationships between leaf N and A_{mass} or dark respiration were weaker when considered on an area basis, as were relationships between leaf P and other traits (Table 3). By also including LMA in analyses, we can quantify the independent effects of leaf structure and nutrient content on A_{mass} and dark respiration^{32,34}. A_{mass} increased with increasing N_{mass} at any given LMA, and decreased with increasing LMA at any given N_{mass} (partial regression coefficients for LMA and leaf N, $P < 0.001$; regression details in Supplementary Information). Similarly, A_{area} increased with increasing N_{area} at any given LMA, and decreased with increasing LMA at any given N_{area} (Fig. 2d and Supplementary Information). That is, both leaf structure and nitrogen concentration affect photosynthetic capacity^{5,6,8,32}. The independent LMA effect is most probably due to leaves with high mass per area having longer diffusion paths from stomata to chloroplasts or greater internal shading of lower chloroplasts, limiting the A_{mass} possible for a given leaf protein content^{14,15,35}. Also, less of the N may be invested in photosynthetic versus non-photosynthetic leaf components in high-LMA species^{36,37}. Similarly, N_{mass} and LMA showed independent effects on R_{mass} (partial regression coefficients both

Table 3 Area basis of bivariate relationships between the six leaf traits

	log LL	log LMA	log A _{area}	log N _{area}	log P _{area}	log R _{area}
log LL		1.71 (1.62, 1.82)	−2.12 (−2.30, −1.96)	−2.36 (−2.19, −2.54)	NA ($P = 0.862$)	NA ($P = 0.211$)
log LMA	0.42 (678)		NA ($P = 0.152$)	1.54 (1.48, 1.59)	1.23 (1.14, 1.32)	1.20 (1.08, 1.34)
log A _{area}	0.13 (512)	0.003 (764)		1.21 (1.13, 1.31)	0.66 (0.58, 0.75)	0.93 (0.83, 1.04)
log N _{area}	0.04 (706)	0.34 (1,958)	0.13 (722)		0.68 (0.64, 0.72)	0.89 (0.80, 0.98)
log P _{area}	0.0002 (202)	0.01 (739)	0.02 (223)	0.35 (750)		1.26 (1.04, 1.54)
log R _{area}	0.01 (217)	0.14 (274)	0.19 (259)	0.34 (267)	0.26 (78)	

Standardized major axis slopes with 95% confidence intervals are given in the upper right section of the matrix (y variable is column 1, x variable in row 1). Coefficients of determination (r^2) and sample sizes are given in the lower left section of the matrix. NA, not applicable; in these cases the correlation was clearly non-significant (P values given in parentheses). While a standardized major axis can still be fitted in such cases, its slope is essentially meaningless. All other relationships were highly significant, $P \ll 0.0001$, with the exception of those between log P_{area} and each of log LMA ($P = 0.029$) and log A_{area} ($P = 0.034$). Further details allowing calculation of predictive regression equations for each pair of traits are given in the Supplementary Information.

$P < 0.001$), although the physiological basis for an independent LMA effect on R_{mass} is less clear than for photosynthetic capacity⁹.

Relating to the weaker and somewhat different pattern of area-based trait relationships, the principal axis of a principal components analysis involving LMA, LL, A_{area} , R_{area} , N_{area} and P_{area} explained a lower proportion of total trait variation than the mass-based equivalent and had a different pattern of trait loadings. This principal axis explained 43% of variation (versus 74% on a mass basis), and largely reflected a spectrum of increasing N_{area} (the trait with the strongest trait loading; Supplementary Information). The five other traits showed positive loadings with the principal axis, with these loadings clearly weaker than in the mass-based analysis. About half the residual variation was explained by a second principal axis, which expressed essentially the same trait correlations as the first axis from the mass-based analysis. Together, these first two axes explained 72% of total trait variation in the area-based data set, marginally less than the first principal axis in the mass-based analysis. Area-based analyses for species grouped by growth form, functional type or biome showed patterns broadly concordant with those across all species, though less clearly than for the mass-based analyses (results not shown).

In summary, the coordination among leaf traits appears to be stronger and simpler on a mass basis than an area basis. This is not because area-basis traits are less varied among species. Rather it is because the LMA–LL spectrum is related to mass-based nutrient concentrations and assimilation and respiration rates in a simpler way than on an area basis.

Climate influence on leaf investment

Plant ecologists have emphasized broad relationships between leaf traits and climate for at least a century. In particular, a general tendency for species inhabiting arid and semi-arid regions to have leathery, high-LMA leaves has been reported^{4,10,38–40}. Building high-LMA leaves needs more investment per unit leaf area. Construction cost per unit leaf mass varies relatively little between species: leaves with high protein content (typically low-LMA leaves) tend to have low concentrations of other expensive compounds such as lipids or lignin, and high concentrations of cheap constituents such as minerals⁴¹. Leaf traits associated with high LMA (for example, thick leaf blade; small, thick-walled cells) have been interpreted as adaptations that allow continued leaf function (or at least postpone leaf death) under very dry conditions, at least in evergreen species.

We characterized sites by annual means of temperature, potential evapotranspiration (PET), vapour pressure deficit (VPD) and solar

irradiance, and MAR. PET, VPD and irradiance were cross-correlated with rainfall and temperature. Against expectations, LMA showed only a very weak relationship with lower rainfall, considered worldwide ($r^2 = 0.002$, $P = 0.032$; 2,370 species from 163 sites). LMA was actually more strongly (positively) correlated with MAT ($r^2 = 0.10$). Worldwide, precipitation is correlated with temperature, cold high-latitude environments typically having low precipitation (Fig. 1). Once variation in MAT was controlled in a multiple regression, LMA did indeed increase as rainfall decreased (Fig. 3). Similarly, LMA was more strongly (positively) correlated with VPD ($r^2 = 0.15$), PET ($r^2 = 0.15$) or site irradiance ($r^2 = 0.18$) than with MAT or site rainfall alone.

Despite the substantial overlap in leaf traits of evergreen and deciduous species, these species groups varied somewhat in their leaf trait–climate relationships. LMA and rainfall showed a strong negative relationship in evergreen shrubs and trees, whereas in deciduous species they were virtually unrelated ($r^2 = 0.22$ versus 0.002). Across all species, LL was positively correlated with PET, VPD, MAT and site irradiance (r^2 values ranging from 0.04 to 0.10). In deciduous trees and shrubs these relationships were consistently stronger (r^2 values ranged from 0.37 to 0.51). That is, LL of deciduous species was shorter at colder sites where the growing season was shorter. But in evergreen shrubs and trees, LL tended to be longer at colder, lower-humidity sites (for example, MAT versus LL, $r^2 = 0.10$; PET versus LL, $r^2 = 0.18$). This is consistent with cold climate vegetation being typically N-limited and demonstrating ‘slow-return’ strategies⁴².

Trait coordination is largely independent of climate

A major aim of the Glopnet collaboration was to obtain enough coverage of climate variation to dissect out effects of climate on relationships between leaf economic traits. There were indeed statistically significant effects of climate. Nevertheless, a major finding from this project is that the influence of climate was, in general, quite modest. How can this be, given that traits such as LMA and LL vary systematically with MAR, MAT and other climate indices? The answer seems to have two elements.

First, much of the total leaf economic variation occurs among co-existing species. The proportion of total variation in LL within sites was 57%, the remaining 43% occurring between sites (variance components analysis). For R_{mass} , the proportion of within-site variation was 67%, for A_{mass} 48%, for N_{mass} 38%, for LMA 36%,

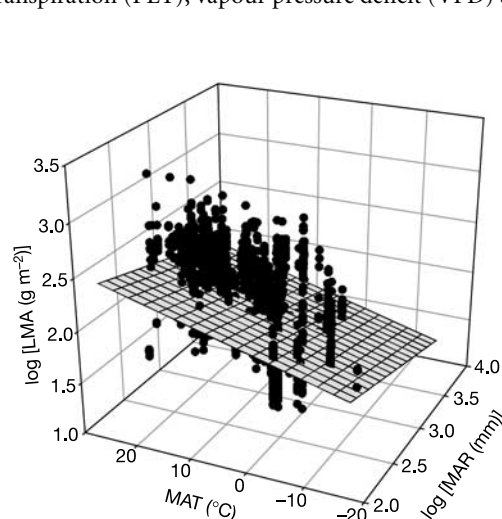


Figure 3 LMA as a function of MAT and MAR at the study sites (data for 2,370 species from 163 sites; rainfall and LMA are $\log_{10}$ -scaled). The coefficients for MAT and log rainfall were highly significant in a multiple regression (both $P < 0.0001$; further details given in Supplementary Information).

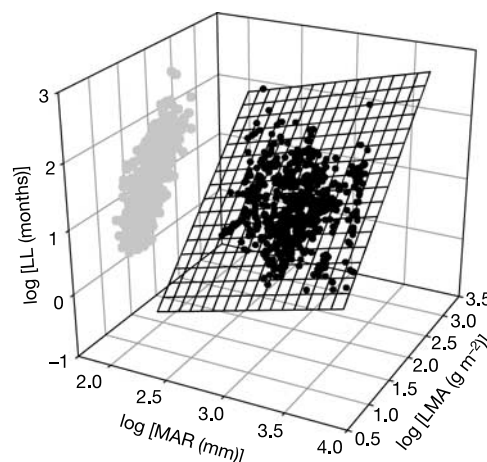


Figure 4 LL as a function of LMA and MAR (all axes are $\log_{10}$ -scaled). When viewed in three dimensions, the two-dimensional LL–LMA cloud of points is spread along a sloping surface. The slope of this surface is steeper in the LMA dimension than in the rainfall dimension, reflecting the higher partial regression coefficient for LMA (1.23 versus 0.47). Both coefficients were highly significant in a multiple regression ($P < 0.0001$; $r^2 = 0.51$; data for 678 species from 51 sites).

for P_{mass} 20%. Similar or higher proportions of within-site variation were seen for area-based traits (40–57% of total variation). Second, leaf traits tend to vary in concert, with the leaf economics spectrum operating similarly in different biomes. Considered across all species, when single or multiple climate variables were included in regressions of LMA on N_{mass} , of A_{mass} on N_{mass} , or of LL on A_{mass} , they added a maximum of 0.05 to the r^2 . Similar results were found for most other bivariate trait relationships, as well as for regression models predicting one leaf trait from two or more other leaf traits as well as climate variables (<15% explanatory power added).

This is not to say that climate does not exert important influences on trait relationships. With such large sample sizes, these effects were still highly statistically significant. Rather, it is an extension of the fact that so much of the total variation in leaf traits is captured by the primary axis of the leaf economics spectrum. Adding climate variables to regressions involving area-based traits also added little explanatory power, for the most part. Exceptions were cases where the bivariate trait relationship was particularly weak to start with (for example, between A_{area} and LMA, R_{area} and LMA, or LL and A_{area}), climate variables added 0.15 to 0.28 to the model r^2 .

LL–LMA relationships shift with climate

The relationship between LMA and LL is important for leaf economics, LMA reflecting the dry-mass cost of deploying new leaf area, and LL representing the duration over which photosynthetic revenue is returned. LL increased with LMA across all species, LMA explaining 42% of variation in LL (Table 1). This relationship showed quite substantial patterning with climate. At a given LMA, LL was shorter at sites of lower rainfall (Fig. 4), meaning that the duration of return from a unit investment in leaf tissue tends to be shorter in such places. In addition, at sites with harsher climate the gain in LL for a given increase in LMA was less. Significant negative interaction terms in the regression models (all $P < 0.001$) indicated decreasing steepness for LL–LMA relationships in hotter, less humid (higher VPD or PET) or higher irradiance environments. The patterning of LL–LMA relationships with rainfall that was seen in our global data set was reported previously for evergreen species from four sites in eastern Australia¹⁷.

Gloptnet results serve to confirm and extend the pattern across several hundred more species, from many more vegetation types. The previous comparisons deliberately contrasted sites with similar MAT. Here, by taking data from many more sites, representing most of the major biomes of the earth, we have been able to identify previously undescribed patterning of LL–LMA relationships with respect to climate, that is, the shallower response of LL to LMA with increasing MAT, VPD, PET and site irradiance.

A possible reason why a given LMA is associated with shorter LL in drier environments involves leaf nitrogen. For evergreen species in semi-arid habitats in eastern Australia¹⁰, the CO_2 concentration inside leaves during photosynthesis was lower at a given stomatal conductance to CO_2 (g_{CO_2}) than for evergreen species at higher rainfall sites. That is, they had higher carbon fixation (A_{area}) at a given stomatal conductance to water per unit leaf area ($g_{\text{H}_2\text{O}}$), because A_{area} is largely the product of CO_2 drawdown and g_{CO_2} , and g_{CO_2} is directly proportional to $g_{\text{H}_2\text{O}}$ (ref. 13). The lower CO_2 concentration inside leaves in the species from semi-arid sites was linked with higher N_{area} , but the high N_{area} corresponded to less robust tissue—a larger proportion of photosynthetic mesophyll—resulting in leaves that were less strong in a biomechanical sense at a given LMA and, it was argued, leading to the observed shorter LL at a given LMA¹⁷.

We do not yet have leaf toughness information or leaf anatomy widely across our global data set, but N_{area} did indeed increase with decreasing site rainfall in evergreen trees and shrubs ($r^2 = 0.18$), as well as across all species (albeit more weakly; $r^2 = 0.04$), CO_2 drawdown increased with increasing N_{area} at a given $g_{\text{H}_2\text{O}}$ (see

Supplementary Information), and A_{area} was higher at a given $g_{\text{H}_2\text{O}}$ at lower rainfall sites (see Supplementary Information), indicating that CO_2 concentration inside leaves tended to be lower at low rainfall sites (correlation $r = 0.44$, $P < 0.0001$). Taken together, these trends can be understood as resulting from the simultaneous optimization of nitrogen- and water-use during photosynthesis^{19,20}. They suggest that similar mechanisms operate in many of the world's vegetation types, as were seen in the eastern Australian species.

Conclusions

We now have wide-ranging and convincing evidence that feasible leaf investment strategies are to a great extent arrayed along a single spectrum, with the same patterning of trait correlations seen globally and in species grouped by growth form, biome or climate. Besides leaf economics, many activities such as seed production, root economics and relations with mycorrhizas are undoubtedly important for plant fitness. But, surely, the broad generality of the relationships we describe suggests that natural selection eventually eradicates leaf investment strategies that are not economically competitive. The leaf economics spectrum reflects a mixture of direct and indirect causal relationships between traits.

For example, the linkage of high A_{mass} with high N_{mass} is in large part the result of a direct causal relationship⁸. Similarly, long LL requires the robustness and low palatability (including chemical defences) associated with high LMA^{17,18}. More indirectly, high A_{mass} tends to be associated with short LL because it requires high N_{mass} and/or low LMA, which increase leaf vulnerability to herbivory and physical hazards, and because high A_{mass} drives fast growth, rapidly shading older leaves, leading them to senesce once their resources become more valuable when transferred to better-lit newer foliage⁴³. The absence of outliers in the trait relationships is particularly striking. Although species vary widely in growth form, life history and niche space occupied, it seems that a mixture of physiological causation and the demands of competitiveness constrain species data points within tightly bounded domains of trait space⁶.

Overall, we found that the effect of climate on leaf trait relationships was modest, although particular trait-pairs showed striking and significant patterns with site climatic properties. Reliable quantification of the global leaf economics spectrum and its relationship to climate should prove valuable in modelling how carbon and nitrogen are tied up and used in plant biomass, and how nutrient fluxes and vegetation boundaries will shift with land-use and climate change. Recent models of vegetation dynamics under global change^{44,45} do make use of the LMA spectrum previously described in ref. 6 for the purpose of grouping species into plant functional types. Recognition of the leaf economics spectrum strengthens and consolidates our knowledge in this regard. At the same time, it is important to note that the spectrum is continuous, rather than divided into distinct categories either by growth form or by habitat. Formulations of plant functional typologies that represent variation as a continuous spectrum, rather than distinct categories, bear promise for the future. □

Methods

Leaf traits

Leaf trait data were compiled from both published and unpublished sources. A data set was considered suitable provided it contained data for at least two of the leaf traits for at least four co-occurring species. Highly artificial vegetation types such as forestry plantations and crop fields were not included. Only site-based data sets were used so that we could reasonably attach climate data. The total data set (Supplementary Information) represented 175 sites and contained 2,548 species–site combinations, consisting of 2,021 species, with 350 occurring at more than one site.

Climate

Climate data were taken from (1) the sites themselves, where known; (2) the nearest weather stations, with temperature data scaled where necessary by an altitudinal lapse rate of 0.6°C per 100 m (ref. 46); (3) a global $0.5^\circ \times 0.5^\circ$ data set of MAT, MAR, vapour pressure and irradiance⁴⁷; (4) a global $0.5^\circ \times 0.5^\circ$ data set of Penman–Monteith PET calculated for

the period 1987–1988 (ref. 48). MAT and rainfall data from the global data set agreed closely with known or nearby-station data (scaled for altitude where necessary), giving us confidence that the vapour pressure and irradiance data were reliable also. Irradiance data were not adjusted for altitude because increases in cloud cover with elevation tend to offset the increase in radiation that would be observed in clear air, although this may not be true for high-elevation sites in arid regions^{46,49}. For high-elevation sites, PET was estimated from a regression equation fitted to all other sites, which considered PET as a function of MAT and annual rainfall ($r^2 = 0.71$). Vapour pressure was scaled for high-elevation sites using an empirical formula expressing the exponential decrease of vapour pressure with altitude⁴⁹. Monthly mean VPD was estimated as the difference between the saturation vapour pressure of air (at the monthly mean temperature) and vapour pressure taken from the global data set. Saturation vapour pressure was calculated using the Tetens formula⁵⁰. Climate variables were averaged or summed (rainfall) across all months of the year, and for those months with mean temperature over 4.99 °C, giving an estimate of climate during the growth season. Results using yearly and growth-season climate indices differed little, so we report results relating to yearly climate averages only.

Data transformation and analysis

For each species at each site, the mean value for each trait was used. Where traits were reported separately for sun leaves and shade leaves, sun-leaf data were used. Similarly, if data were presented separately for recently matured and old leaves, that of the recently matured leaves were chosen. Leaf traits were approximately log-normally distributed across the data set, as were rainfall and VPD. Accordingly, these variables were log₁₀-transformed before analysis. MAT, PET and solar radiation were left untransformed because their distribution across sites was approximately normal.

Standardized major axis slopes⁵⁴ with 95% confidence intervals were fitted to bivariate trait relationships because our aim was to describe the best-fit lines or central axes of these 'scaling' relationships. Ordinary multiple regression was used for analyses exploring the additional predictive power of climate variables on leaf trait relationships. Regressions were first run including interaction terms between climate variables and leaf traits; where the interaction was non-significant ($P < 0.05$), we re-ran models with main effects only. In several cases, principal components analyses run on data subsets defined by growth form or plant functional type had to be re-run with one or more leaf traits removed, owing to insufficient data. These cases are indicated in Table 2. Variance component analyses were based on the decomposition of analysis of variance (ANOVA) type I sums of squares. Principal components analyses, variance components and regression analyses were run in SPSS for Windows version 11.01.

Received 15 July 2003; accepted 10 February 2004; doi:10.1038/nature02403.

- Bloom, A. J., Chapin, F. S. & Mooney, H. A. Resource limitation in plants—an economic analogy. *Annu. Rev. Ecol. Syst.* **16**, 363–392 (1985).
- Orians, G. H. & Solbrig, O. T. A cost-income model of leaves and roots with special reference to arid and semiarid area. *Am. Nat.* **111**, 677–690 (1977).
- Givnish, T. J. (ed.) *On the Economy of Plant Form and Function* (Cambridge Univ. Press, Cambridge/New York, 1986).
- Niinemets, U. Global-scale climatic controls of leaf dry mass per area, density, and thickness in trees and shrubs. *Ecology* **82**, 453–469 (2001).
- Niinemets, U. Components of leaf dry mass per area—thickness and density—alter leaf photosynthetic capacity in reverse directions in woody plants. *New Phytol.* **144**, 35–47 (1999).
- Reich, P. B., Walters, M. B. & Ellsworth, D. S. From tropics to tundra: global convergence in plant functioning. *Proc. Natl Acad. Sci. USA* **94**, 13730–13734 (1997).
- Reich, P. B. *et al.* Generality of leaf trait relationships: a test across six biomes. *Ecology* **80**, 1955–1969 (1999).
- Field, C. & Mooney, H. A. in *On the Economy of Plant Form and Function* (ed. Givnish, T. J.) 25–55 (Cambridge Univ. Press, Cambridge, 1986).
- Reich, P. B. *et al.* Relationships of leaf dark respiration to leaf nitrogen, specific leaf area and leaf life-span—a test across biomes and functional groups. *Oecologia* **114**, 471–482 (1998).
- Wright, I. J., Reich, P. B. & Westoby, M. Strategy-shifts in leaf physiology, structure and nutrient content between species of high and low rainfall, and high and low nutrient habitats. *Funct. Ecol.* **15**, 423–434 (2001).
- Grime, J. P. *et al.* Integrated screening validates primary axes of specialisation in plants. *Oikos* **79**, 259–281 (1997).
- Whittaker, R. *Communities and Ecosystems* 2nd edn, 167 (MacMillan, New York, 1975).
- Lambers, H., Chapin, F. S. & Pons, T. L. *Plant Physiological Ecology* (Springer, New York, 1998).
- Parkhurst, D. F. Diffusion of CO₂ and other gases inside leaves. *New Phytol.* **126**, 449–479 (1994).
- Evans, J. R. & Loreto, F. in *Photosynthesis: Physiology and Metabolism* (eds Leegood, R. C., Sharkey, T. D. & von Caemmerer, S.) 321–351 (Kluwer Academic, Dordrecht, 2000).
- Cannell, M. G. R. & Thornley, J. H. M. Modelling the components of plant respiration: Some guiding principles. *Ann. Bot.* **85**, 45–54 (2000).
- Wright, I. J. & Westoby, M. Leaves at low versus high rainfall: coordination of structure, lifespan and physiology. *New Phytol.* **155**, 403–416 (2002).
- Coley, P. D. Herbivory and defensive characteristics of tree species in a lowland tropical forest. *Ecol. Monogr.* **53**, 209–233 (1983).
- Wright, I. J., Reich, P. B. & Westoby, M. Least-cost input mixtures of water and nitrogen for photosynthesis. *Am. Nat.* **161**, 98–111 (2003).
- Farquhar, G. D., Buckley, T. N. & Miller, J. M. Optimal stomatal control in relation to leaf area and nitrogen content. *Silva Fennica* **36**, 625–637 (2002).

- Reich, P. B., Walters, M. B. & Ellsworth, D. S. Leaf life-span in relation to leaf, plant, and stand characteristics among diverse ecosystems. *Ecol. Monogr.* **62**, 365–392 (1992).
- Lambers, H. & Poorter, H. Inherent variation in growth rate between higher plants: a search for ecological causes and consequences. *Adv. Ecol. Res.* **23**, 187–261 (1992).
- Wright, I. J. & Westoby, M. Nutrient concentration, resorption and lifespan: leaf traits of Australian sclerophyll species. *Funct. Ecol.* **17**, 10–19 (2003).
- Sokal, R. R. & Rohlf, F. J. *Biometry: the Principles and Practice of Statistics in Biological Research* (W. H. Freeman, New York, 1995).
- Koerselman, W. & Meuleman, A. F. M. The vegetation N/P ratio—a new tool to detect the nature of nutrient limitation. *J. Appl. Ecol.* **33**, 1441–1450 (1996).
- Sterner, R. W. & Elser, J. J. *Ecological Stoichiometry. The Biology of Elements from Molecules to the Biosphere* (Princeton Univ. Press, Princeton/Oxford, 2002).
- Westoby, M., Warton, D. & Reich, P. B. The time value of leaf area. *Am. Nat.* **155**, 649–656 (2000).
- Grime, J. P. in *Exploitation of Environmental Heterogeneity in Plants* (eds Caldwell, M. M. & Pearcy, R. W.) 1–18 (Academic, San Diego, 1994).
- Harper, J. L. The value of a leaf. *Oecologia* **80**, 53–58 (1989).
- Walters, M. B. & Reich, P. B. Seed size, nitrogen supply, and growth rate affect tree seedling survival in deep shade. *Ecology* **81**, 1887–1901 (2000).
- Reich, P. B. & Walters, M. B. Photosynthesis-nitrogen relations in Amazonian tree species. 2. Variation in nitrogen vis-a-vis specific leaf area influences mass-based and area-based expressions. *Oecologia* **97**, 73–81 (1994).
- Reich, P. B., Ellsworth, D. S. & Walters, M. B. Leaf structure (specific leaf area) modulates photosynthesis-nitrogen relations: evidence from within and across species and functional groups. *Funct. Ecol.* **12**, 948–958 (1998).
- Schulze, E.-D., Kelliher, F. M., Körner, C., Lloyd, J. & Leuning, R. Relationships among maximum stomatal conductance, ecosystem surface conductance, carbon assimilation rate, and plant nitrogen nutrition: a global scaling exercise. *Annu. Rev. Ecol. Syst.* **25**, 629–660 (1994).
- Peterson, A. G. CMEAL participants. Reconciling the apparent difference between mass- and area-based expressions of the photosynthesis-nitrogen relationship. *Oecologia* **118**, 114–150 (1999).
- Green, D. S. & Kruger, E. L. Light-mediated constraints on leaf function correlate with leaf structure among deciduous and evergreen tree species. *Tree Physiol.* **21**, 1341–1346 (2001).
- Poorter, H. & Evans, J. R. Photosynthetic nitrogen-use efficiency of species that differ inherently in specific leaf area. *Oecologia* **116**, 26–37 (1998).
- Hikosaka, K., Hanba, Y. T., Hirose, T. & Terashima, I. Photosynthetic nitrogen-use efficiency in leaves of woody and herbaceous species. *Funct. Ecol.* **12**, 896–905 (1998).
- Maximov, N. A. *The Plant in Relation to Water. A Study of the Physiological Basis of Drought Resistance* (Allen & Unwin, London, 1929).
- Schulze, E.-D. *et al.* Carbon and nitrogen isotope discrimination and nitrogen nutrition of trees along a rainfall gradient in northern Australia. *Aust. J. Plant Physiol.* **25**, 413–425 (1998).
- Fonseca, C. R., Overton, J. M., Collins, B. & Westoby, M. Shifts in trait combinations along rainfall and phosphorus gradients. *J. Ecol.* **88**, 964–977 (2000).
- Villar, R. & Merino, J. Comparison of leaf construction costs in woody species with differing leaf lifespans in contrasting ecosystems. *New Phytol.* **151**, 213–226 (2001).
- Oleksyn, J., Reich, P. B., Zytowski, R., Karolewski, P. & Tjoelker, M. G. Nutrient conservation increases with latitude of origin in European *Pinus sylvestris* populations. *Oecologia* **136**, 220–235 (2003).
- Ackerly, D. D. & Bazzaz, F. A. Leaf dynamics, self-shading and carbon gain in seedlings of a tropical pioneer tree. *Oecologia* **101**, 289–298 (1995).
- Bonan, G. B., Levis, S., Kergoat, L. & Oleson, K. W. Landscapes as patches of plant functional types: an integrating concept for climate and ecosystem models. *Glob. Biogeochem. Cycles* **16**, doi:10.1029/2000GB001360 (2002).
- Moorcroft, P. R., Hurr, G. C. & Pacala, S. W. A method for scaling vegetation dynamics: The ecosystem demography model (ED). *Ecol. Monogr.* **71**, 557–585 (2001).
- Körner, C. *Alpine Plant Life: Functional Plant Ecology of High Mountain Ecosystems* Ch. 3 (Springer, Berlin, 1999).
- New, M., Hulme, M. & Jones, P. Representing twentieth-century space-time climate variability. Part I: Development of a 1961–90 mean monthly terrestrial climatology. *J. Clim.* **12**, 829–856 (1999).
- Choudhury, B. J. Global pattern of potential evaporation calculated from the Penman-Monteith equation using satellite and assimilated data. *Remote Sens. Environ.* **61**, 64–81 (1997).
- Barry, R. G. *Mountain Weather and Climate* 24–29 (Methuen, London/New York, 1981).
- Campbell, G. S. & Norman, J. M. *An Introduction to Environmental Biophysics* (Springer, New York, 1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements All authors contributed (mostly unpublished) data and intellectual input to the project. The first three authors took the lead in the organization, analysis and writing-up of this work, and contributed 8% of the data, with 223 species from 11 sites. We also thank the many other researchers who provided additional information about their study sites and published data.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to I.W. (iwright@rna.bio.mq.edu.au).

Prediction of a global climate change on Jupiter

Philip S. Marcus

Department of Mechanical Engineering, University of California, Berkeley, California 94720, USA

Jupiter's atmosphere, as observed in the 1979 Voyager space craft images, is characterized by 12 zonal jet streams and about 80 vortices, the largest of which are the Great Red Spot and three White Ovals that had formed¹ in the 1930s. The Great Red Spot has been observed² continuously since 1665 and, given the dynamical similarities between the Great Red Spot and the White Ovals, the disappearance^{3,4} of two White Ovals in 1997–2000 was unexpected. Their longevity and sudden demise has been explained⁵ however, by the trapping of anticyclonic vortices in the troughs of Rossby waves, forcing them to merge. Here I propose that the disappearance of the White Ovals was not an isolated event, but part of a recurring climate cycle which will cause most of Jupiter's vortices to disappear within the next decade. In my numerical simulations, the loss of the vortices results in a global temperature change of about 10 K, which destabilizes the atmosphere and thereby leads to the formation of new vortices. After formation, the large vortices are eroded by turbulence over a time of ~ 60 years—consistent with observations of the White Ovals—until they disappear and the cycle begins again.

Long-lived cyclones are essential to my climate cycle, but many investigators do not believe they exist⁶. I therefore begin by arguing for their existence. The argument^{7,8} against long-lived cyclones is that their tangled, filamentary clouds are inconsistent with streamlines of coherent vortices. However, inferring longevity from cloud morphology is difficult. Moreover, cyclonically rotating clouds are currently visible in the same regions where Voyager observed them in 1979, suggesting that the underlying cyclones are long-lived. Figure 1, which compares observed and numerically simulated clouds, attacks the morphology argument. (The Methods section describes these simulations.) The statistically steady equilibrium has two, staggered, opposite-signed rows of vortices called a Kármán vortex street. All the vortices last indefinitely, so the cyclones are long-lived. Because jovian vortices⁹ are patches of nearly uniform potential vorticity q , a vortex is labelled an 'anticyclone' if its $q > 0$ (Fig. 1b). (I adopt a southern hemisphere reference, so anticyclones rotate anticlockwise, but in both hemispheres the anticyclones spin opposite to the planet's rotation.) In Fig. 1c, the centre of an anticyclone has vorticity ω that is also anticyclonic ($\omega > 0$) but is surrounded by a band of cyclonic ω . 'Cyclones' have the opposite arrangement. Although the velocities of cyclones and anticyclones in Fig. 1 are mirror images, their simulated clouds are different because clouds act as particle tracers and in turbulent flows particle paths are not coincident with streamlines. My simulations create clouds where $\omega > 0$ and melt them where $\omega < 0$. Ice-forming and -melting locations differ for cyclones and anticyclones. Figure 1 shows that observed and simulated clouds are similar, that clouds of cyclones and anticyclones are dissimilar, and (most importantly) that twisted, filamentary clouds are consistent with long-lived cyclones.

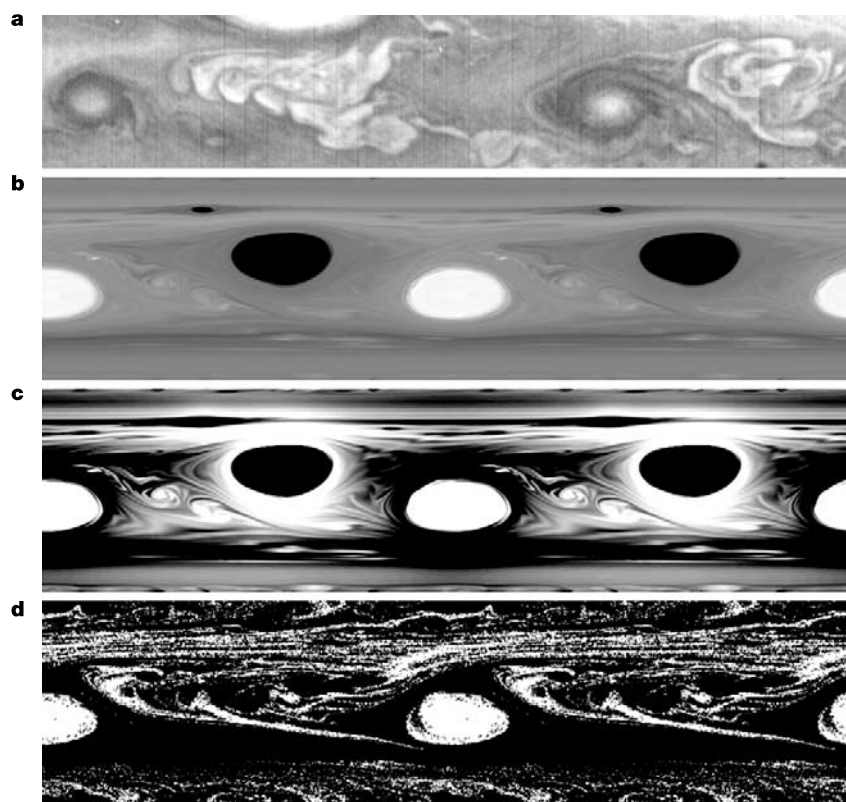


Figure 1 Observed and simulated jovian clouds. **a**, Voyager mosaic showing 2 of the 12 cyclone/anticyclone pairs at 41°S . The clouds of 'cyclones' are large, tangled and filamentary, whereas those of 'anticyclones' are bright, 2:1 ovals surrounded by dark rings. **b**, Simulated potential vorticity q , with cyclonic q shown black and anticyclonic q white. **c**, ω of the same flow in **b** with cyclonic ($\omega < 0$) shown black and anticyclonic ω white. **d**, Simulated clouds of the flow in **b**, where white pixels represent NH_3 ice. For

anticyclones: ice is created in the vortex interior where $\omega > 0$ and there is weak upwelling; ice is rapidly melted in the ring around the vortex where $\omega < 0$; the clouds are ovals (coincident with regions where $\omega > 0$) surrounded by dark rings (coincident with $\omega < 0$). For cyclones: turbulence spreads and twists the clouds created at the peripheral rings (where $\omega > 0$); the turbulence fills the interiors (where $\omega < 0$) with cloud filaments before they melt.

My climate cycle has five stages, the longest of which is the ‘Kármán vortex streets’ stage. I believe that in the current cycle, this stage began in about 1940 and is now ending. With the exception of the Great Red Spot (GRS), all long-lived jovian vortices occur in opposite-signed pairs in streets. Anticyclones are longitudinally staggered with cyclones, so that like-signed vortices are never adjacent (Figs 1 and 2). A westward-going jet stream separates the street’s two rows. The vortices drift and oscillate in longitude with velocities $\sim 3 \text{ m s}^{-1}$ (jet streams and vortex winds are $\sim 100 \text{ m s}^{-1}$). My calculations⁹ showed that an initial row of anticyclones, with no neighbouring cyclones, was unstable: the vortices approached each other, and they all merged. The only way I could maintain them in a row was by arranging them in a vortex street. I found¹⁰ that cyclones and anticyclones repel in this configuration. If, for any reason, two same-signed vortices become adjacent, they rapidly merge. To push an anticyclone past a neighbouring cyclone and make two anticyclones longitudinally adjacent (that is, with no intervening cyclone) requires a perturbation velocity V_{crit} exceeding $\sim 30 \text{ m s}^{-1}$. A new analysis in the Methods section agrees with this finding.

As vortex streets are essential to the climate cycle proposed here, the GRS’s existence outside a street (that is, its uniqueness at its latitude) needs clarification. Jovian vortices are robust because strong Coriolis forces make the atmospheric flow nearly two-dimensional, an environment where vortices thrive. (Three-dimensional flow destroys vortices.) The GRS cannot be part of a street because the street’s cyclones would need to be north of the westward jet stream at 20°S , and no cyclones (not even transients) lie between 20°S and the equator. Their absence is consistent with the flow’s more three-dimensional appearance there (presumably due to the weak Coriolis force). Ephemeral anticyclones often appear at the same latitude as the GRS but, because there are no nearby cyclones to repel them, they merge with it, leaving the GRS as the only long-lived vortex there.

As $V_{\text{crit}} \approx 30 \text{ m s}^{-1}$ is needed to disrupt a vortex street and the available perturbations from turbulence and longitudinal oscillations are only $\sim 3 \text{ m s}^{-1}$, it was surprising in 1994 when two White Ovals became closely spaced, drifted together for 3 years and then merged. These events marked the cycle’s ‘Trapping and vortex merging’ stage. Experiments¹¹ and calculations¹² show that the trough of a Rossby wave (on an eastward jet stream) can trap vortices, making them and the trough drift together⁵ (Fig. 3). Vortices are trapped because anticyclones are repelled by the cyclonic outside edge of the trough as if it were a cyclone.

Varying the size or area A_C of the trapped cyclone while holding all other parameters to be those of the White Ovals, we found⁵ three critical areas: $A_C^{\text{escape}} > A_C^{\text{trap}} > A_C^{\text{merge}}$. Untrapped vortices get trapped when $A_C < A_C^{\text{trap}}$; when $A_C > A_C^{\text{escape}}$, initially trapped

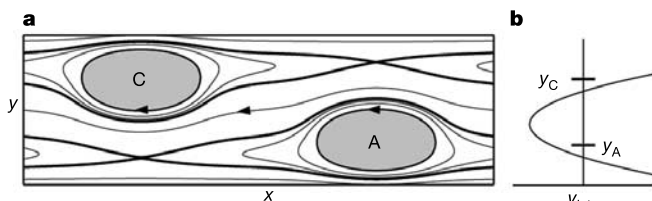


Figure 2 Dynamics of two vortices in a Kármán street. **a**, Streamlines of one period (in the east-west or x direction) of an infinitely long vortex street with cyclone C and anticyclone A. Jovian anticyclones (cyclones) are on the poleward (equatorial) side of the westward jet. There are three pairs of opposite-signed vortices straddling the jet at 34°S (in which the anticyclones are the White Ovals), 12 at 41°S , and so on. The outermost closed streamlines (OCS) which connect the stagnation points (where streamlines cross) are heavy lines. Outside the OCS the streamlines are open. To represent the White Ovals, I would need to increase the spacings between the vortices by a factor of ~ 6 while holding the sizes of the vortices fixed. **b**, The velocity of the jet stream as a function of y . For the numerical experiment described in the Methods section, the initial locations of the latitudes of the centres of the anticyclone and cyclone are y_A and y_C .

vortices escape the trough; and when $A_C < A_C^{\text{merge}}$, the repulsion is so weak that the two trapped anticyclones become adjacent and merge. When $A_C^{\text{trap}} > A_C > A_C^{\text{merge}}$, the trough ‘squeezes’ trapped vortices together⁵, causing V_{crit} to decrease from $\sim 30 \text{ m s}^{-1}$ to $\sim 3 \text{ m s}^{-1}$. We believe that two White Ovals (and a cyclone) drifted together (that is, were trapped) in 1994–97 because A_C became less than A_C^{trap} . Just before their merger, they collided with an untrapped cyclone (Fig. 3). This created a perturbation greater than $V_{\text{crit}} \approx 3 \text{ m s}^{-1}$, and the Ovals merged. We showed⁵ that if A_C continues to decay, vortices in a street become trapped and merge until only one vortex pair survives.

The Methods section shows that vortex mergers lead to ‘Global temperature changes’, the cycle’s next stage, in which the temperature T near the equator (poles) rises (falls) by $\sim 10 \text{ K}$. Currently the weather layer (containing the clouds and vortices) is nearly isothermal in latitude. This is surprising and not understood. A balance between cooling via blackbody radiation and heating from the Sun (a function of latitude) and internal sources would make the poles $\sim 30 \text{ K}$ cooler than the equator¹³. Most solar heat is absorbed in, or just below, the weather layer, so deep convection cannot make its T uniform¹⁴. Consistent with theory¹⁵ my calculations show that when there are several vortices per westward jet, the velocity $\mathbf{v}$ is chaotic and chaotic mixing of T makes the layer isothermal. An exception is that the T within and exterior to the vortices do not mix. The strong gradient of q circumscribing each vortex acts as a barrier to transport¹⁶. Chaotic mixing explains not only why the layer is isothermal but also why cyclones (anticyclones) remain warm (cool). My calculated T is insensitive to details of $\mathbf{v}$ as long as $\mathbf{v}$ is chaotic and $D/||\mathbf{v}|| \ll \tau_{\text{rad}}$, that is, the advective time (a few days) is much less than the radiative time ($\tau_{\text{rad}} \approx 7$ years), where D is a vortex diameter. After the vortices in the streets merge, the chaos and mixing diminish, T relaxes toward its radiative equilibrium, a large latitudinal gradient in T forms, and at subtropical latitudes the T at the layer’s bottom (where the solar heat is absorbed) increases by $\delta T \approx 6 \text{ K}$. (Note that vortices create high

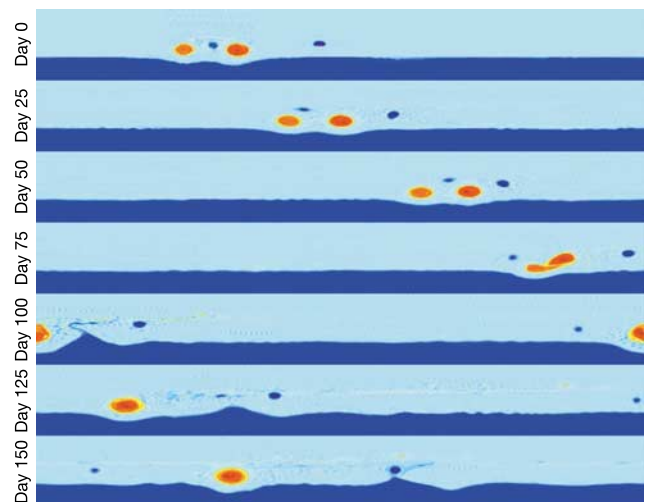


Figure 3 Frames of q from a simulation movie. (Available as Supplementary Information.) The domain is periodic in x . Initially, two anticyclones (red) and one cyclone (blue) are trapped in the trough of a Rossby wave (interface between light and dark blue), and there is also a small cyclone 25,000 km east of the trough. The initial conditions mimic the White Ovals before their merger⁵. For the first 60 days, the trapped vortices drift eastward at $\sim 1 \text{ m s}^{-1}$ and are closely spaced inside the trough. The centre cyclone oscillates between the outer anticyclones. (Without the trough, they spread apart at speeds of $\sim 5 \text{ m s}^{-1}$.) The untrapped cyclone initially moves west, but after day ~ 60 moves east because it is repelled by the vortices in the trough. The repulsion perturbs the central cyclone, allowing it to pass to the west of the anticyclone on its westward side at day ~ 60 . The anticyclones are then adjacent and merge (days 75–100). The remaining cyclones merge ~ 1 year later.

clouds that absorb and reflect heat, and thereby cool the bottom of the weather layer. Thus the loss of vortices increases δT for a second, independent reason.) Currently the weather layer is not strongly stratified. The Methods section shows that a ~ 6 K warming is not enough to destabilize the layer to convection, but is strong enough to decrease the Rossby deformation radius L_r by 30% which will 'Destabilize jet streams and form new vortices'. This is the cycle's next stage. Calculations show that after the vortices merge, leaving only one pair per westward jet stream, the flow is stable if L_r is held fixed. However, when L_r is reduced in the calculations to be in accord with temperature changes, within a few months waves form on the jet streams. They grow, break and roll-up into new vortices, in agreement with others' calculations¹⁷ of baroclinic jet instabilities. I believe that the break-up of the band of clouds in 1938–39 that formed the White Ovals is an example of this instability. A necessary condition for linear instability¹⁸ of a steady jet is that its $q(y)$ has extrema. Extrema are created when L_r decreases: equation (1) in Methods shows that if L_r^2 decreases by half while $\mathbf{v}$ remains fixed, then q changes from $q_{\text{jet}}^{\text{old}}$ to $q_{\text{jet}}^{\text{new}}(y) \equiv q_{\text{jet}}^{\text{old}}(y) - \psi_{\text{jet}}(y)/[L_r^{\text{old}}]^2$, which generally has large extrema because the velocity of the jet stream $v_{\text{jet}}(y)$ (and therefore the stream function of the jet stream $\psi_{\text{jet}}(y)$) is oscillatory.

New vortices rapidly shrink until they look like the vortex streets in the cycle's first stage. The newly formed White Ovals shrank¹ quickly from $\sim 90,000$ km to $\sim 20,000$ km. From 1950 to 2000, they slowly shrank to $\sim 10,000$ km. Calculations show that the 'Slow erosion of vortex streets' stage continues for ~ 60 years. This is the cycle's final stage. Figure 2 explains the shrinkage. Closed streamlines surround vortices. The outermost closed streamline (OCS) contains stagnation points. When a vortex forms that is larger than its OCS, the q outside the OCS quickly pulls away until the vortex lies within its OCS. This was the likely cause of the Ovals' initial rapid shrinkage. A vortex random-walks around its equilibrium at the centre of its OCS owing to turbulence. When its edge bumps the OCS, its outer q is carried to a stagnation point where it is stripped away. The stronger the turbulence, the smaller the surviving vortex¹⁹. Turbulent erosion is analogous to carrying a full cup of coffee: with steady (laminar) hands, the cup remains filled to its brim (OCS). With unsteady hands, the mean surface of the coffee is below the brim. The distance from the brim to the surface increases with increasing unsteadiness, so the volume of coffee (vortex area) decreases with increasing unsteadiness (turbulence). Turbulent erosion is more complex for vortex streets because the OCSs are highly elongated, but it explains observations that show that the Ovals' slow shrinkages between 1950 and 2000 were temporally correlated with their reversals in direction¹. Before 1994, the Ovals and cyclones made long oscillations in longitude with periods of ~ 15 years. Their erosion had nearly stopped because they were inside and generally far from their OCSs. However, every half-period, an Oval and cyclone approached each other (and a stagnation point on the OCS) and reversed directions. I argue that during each encounter they eroded. Calculations show that ~ 4 periods (~ 60 years for the Ovals) are needed for A_C to become less than A_C^{trap} . Turbulent erosion is the slowest part of the cycle, and sets its period of $(60 + \tau_{\text{rad}}) \approx 70$ years.

I predict an imminent, dramatic change in the jovian climate as part of a ~ 70 -year cycle. Calculations show that cyclones are required to maintain rows of anticyclones. The cyclones' decay inevitably destroys anticyclones. I believe that the disappearance of two White Ovals at 34° S in the late 1990s was due to the decay of cyclones, and that it was part of a recurring cycle—the disappearance and re-appearance¹ of anticyclones at 34° S in the 1930s was part of the previous cycle. For a cycle to exist, feedback is needed between the disappearance of anticyclones and the dynamics that regenerates them. An obvious feedback is heat transport. Calculations show that chaotic mixing of heat owing to the chaos produced by multiple vortex pairs in a street makes the weather layer isothermal. Destruction of the vortices reduces the chaos and decreases the

mixing. The jovian jet streams are easily destabilized by small temperature changes. Generically, jet streams go unstable via growing waves that break, roll up and create rows of vortices^{10,17,18,20,21}. Therefore, I expect vortices to re-emerge from destabilized jet streams. The true tests of predictions are observations, and there are many to look for within the next decade: the bunching together of vortices (especially at 41° S) followed by their merging, cooling of the poles and warming of the equator, and instabilities of the jet streams indicated by the break-up of circumferential bands of clouds into new 'spots'. $\square$

Methods

Calculations use the quasigeostrophic²² (QG) equation for the potential vorticity q :

$$q \equiv \nabla^2 \psi - \psi/L_r^2 + \beta y; \quad dq/dt \equiv (\partial/\partial t + \mathbf{v} \cdot \nabla)q = 0 \quad (1)$$

with stream function ψ , velocity $\mathbf{v} \equiv \hat{\mathbf{z}} \times \nabla \psi$, vertical unit vector $\hat{\mathbf{z}}$, Coriolis parameter β , north–south coordinate y and Rossby deformation radius L_r . In Fig. 1, the initial distributions of cyclonic and anticyclonic q are mirror images, and owing to the symmetry of equation (1) remain so. I decompose $\mathbf{v}$ into a jet stream component $v_{\text{jet}}(y)\hat{\mathbf{x}}$ plus a remainder consisting of vortices and turbulence $\mathbf{v}'(x, y, t)$, and define the vorticity $\omega \equiv \hat{\mathbf{z}} \cdot (\nabla \times \mathbf{v})$. In the upper weather layer, anticyclonic (cyclonic) ω creates upward (downward) flow¹⁰. In a subadiabatic atmosphere, upwelling cools the fluid, so I simulate the formation of NH_3 ice clouds by adding white pixels (that advect with $\mathbf{v}$) at a rate proportional to the local anticyclonic ω (and upwelling). I melt ice (that is, destroy white pixels) in regions with cyclonic ω . The vortex centres have weak $|\omega|$ and are surrounded by small-area rings (with widths $\sim L_r$) with opposite-signed ω (Fig. 1). The rings' $|\omega|$ are large; the circulation of each vortex $\oint \omega dA$ is zero—as it must be for a compact QG vortex. The advective timescale of turbulent transport is between the slow timescale of melting/creation in the small- $|\omega|$ interior and the fast timescale of the melting/creation in the large- $|\omega|$ ring. For an anticyclone: ice is created and well-mixed in the vortex interior before it advects out, so the interior is uniformly bright. When advected out of the centre, the ice melts quickly leaving the surrounding ring ice-free and dark. For a cyclone: ice is rapidly created in the ring and advects to its surroundings. It remains a long time in the interior where it is stretched into filaments because the interior's melting time is long.

Figure 2 shows why opposite-signed vortices in a Kármán vortex street straddling a westward jet stream repel. Vortices advect with the ambient $\mathbf{v}$. The velocity due to a QG vortex falls off exponentially¹⁰ at distances greater than L_r , so for widely separated vortices, $v_{\text{jet}}(y)$ dominates. Let a cyclone approach an anticyclone from the west. Initially, they advect with velocities $v_{\text{jet}}(y_C)$ and $v_{\text{jet}}(y_A)$, where y_C and y_A are the latitudes of the centres of the cyclone and anticyclone. Figure 2 is in the reference frame with $v_{\text{jet}}(y_C) + v_{\text{jet}}(y_A) = 0$, so initially $v_{\text{jet}}(y_C) > 0$ and $v_{\text{jet}}(y_A) < 0$, and the vortices approach each other. When the vortex separation becomes less than L_r , the anticyclone (cyclone) causes the cyclone (anticyclone) to orbit anticlockwise (clockwise) around it. This pushes both vortices south. The shear of the westward jet stream, $\sigma(y) \equiv -dv_{\text{jet}}/dy$, is negative (positive) on its northern (southern) side, so as the vortices move south, $v_{\text{jet}}(y_C)$ becomes more negative and $v_{\text{jet}}(y_A)$ more positive. The vortices repel. We showed numerically² that a cyclone and anticyclone would need to approach initially at velocities greater than a critical value $V_{\text{crit}} \approx 30 \text{ m s}^{-1}$ to not be completely repelled and to pass by each other (resulting in adjacent like-signed vortices). I can obtain this same value of V_{crit} analytically. An anticyclone's east–west velocity v_A is primarily due to v_{jet} , so a change in its latitude Δy_A changes v_A by:

$$\Delta v_A \approx \Delta y_A (dv_{\text{jet}}/dy) = -\sigma(y_A) \Delta y_A \quad (2)$$

the anticyclone's Δy_A in Fig. 2 is due to the cyclone's velocity, which, using the Biot–Savart law (assuming the vortex separation is $\sim L_r$), has strength $A_C q_C / 2\pi L$ where L is the distance between the vortices and A_C and q_C are the area and q of the cyclone. Therefore, $|\Delta y_A| \approx A_C q_C \Delta t / 2\pi L$, where Δt is the duration of the vortex/vortex encounter. Thus

$$|\Delta v_A| = |\sigma(y_A) A_C q_C \Delta t / 2\pi L| = |\sigma(y_A) A_C q_C / 2\pi v_A| \quad (3)$$

where I used $L/\Delta t \approx |v_A|$. In order for the vortices to repel, $|v_A| < |\Delta v_A|$, or using equation (3):

$$|v_A| < \sqrt{\sigma(y_A) A_C q_C / 2\pi} \equiv V_{\text{crit}} \quad (4)$$

Thus, $V_{\text{crit}} \equiv \sqrt{\sigma(y_A) A_C q_C / 2\pi} = 26 \text{ m s}^{-1}$, where we used parameter values of the White Ovals⁵.

To find changes in the jovian T , I solved numerically the advection equation²³ for T in a thin weather layer differentially heated by the Sun, uniformly heated from below and radiatively cooled. (I used an albedo that gives an average (over the entire atmosphere) incident solar flux of 8.4 W m^{-2} , and I used an internal flux of 5.7 W m^{-2} .) The advection velocity $\mathbf{v}$ is computed from equation (1) and changes with the cycle's stages. The radiative time $\tau_{\text{rad}} \equiv c_p H / 4aT^3 \approx 7$ years, with heat capacity c_p , density ρ , layer thickness $H = 10$ km, and Stefan's constant a . The layer is not strongly stratified; the temperature difference δT between its bottom and top is not much smaller than the value of a neutrally stratified layer $\delta T|_b \equiv gH/c_p = 47$ K, where g is the acceleration due to gravity. Although the jovian δT cannot be measured directly, $L_r(\theta) \equiv \sqrt{gH(\delta T|_b - \delta T)/T/2\Omega \sin \theta}$ is well known¹⁰, where $2\pi/\Omega$ is the jovian period and θ the latitude. This expression gives $\delta T = 35$ K. Heating the bottom of the weather layer by more than $(\delta T|_b - \delta T) = 12$ K would destabilize it to convection; heating it by only ~ 6 K decreases $L_r(\theta)^2$ by half.

Received 21 January; accepted 9 March 2004; doi:10.1038/nature02470.

1. Rogers, J. H. *The Giant Planet Jupiter* (Cambridge Univ. Press, Cambridge, 1995).
2. Hook, R. A spot in one of the belts of Jupiter. *Phil. Trans.* **1**, 3 (1665).

3. Sánchez-Lavega, A. *et al.* Interactions of Jovian White Ovals BC and DE in 1998 from Earth-based observations in the visual range. *Icarus* **142**, 116–124 (1999).
4. Sánchez-Lavega, A. *et al.* The merger of two giant anticyclones in the atmosphere of Jupiter. *Icarus* **149**, 491–495 (2001).
5. Youssef, A. & Marcus, P. S. The dynamics of jovian white ovals from formation to merger. *Icarus* **162**, 74–93 (2003).
6. MacLow, M.-M. & Ingersoll, A. P. Merging of vortices in the atmosphere of Jupiter: an analysis of Voyager images. *Icarus* **65**, 353–369 (1986).
7. Neelin, M. Rossby solitary vortices, on giant planets and in the laboratory. *Chaos* **4**, 187–202 (1994).
8. Sutyrin, G. Long lived planetary vortices and their evolution: Conservative intermediate geostrophic model. *Chaos* **4**, 203–212 (1994).
9. Marcus, P. S. Numerical simulations of Jupiter's Great Red Spot. *Nature* **331**, 693–696 (1988).
10. Marcus, P. S. Jupiter's Great Red Spot and other vortices. *Annu. Rev. Astron. Astrophys.* **31**, 523–573 (1993).
11. Solomon, T. H., Holloway, W. J. & Swinney, H. L. Shear flow instabilities and Rossby waves in barotropic flow in a rotating annulus. *Phys. Fluids* **5**, 1971–1982 (1993).
12. Marcus, P. S. & Lee, C. A model for eastward and westward jets in laboratory experiments and planetary atmospheres. *Phys. Fluids* **10**, 1474–1489 (1998).
13. Flasar, F. M. Global dynamics and thermal structure of Jupiter's atmosphere. *Icarus* **65**, 280–303 (1986).
14. Gierasch, P. J. Radiative-convective latitudinal gradients for Jupiter and Saturn models with a radiative zone. *Icarus* **142**, 148–154 (1999).
15. Ottino, J. M. *The Kinematics of Mixing: Stretching, Chaos, and Transport* (Cambridge Univ. Press, Cambridge, 1989).
16. Solomon, T. H., Weeks, E. R. & Swinney, H. L. Chaotic advection in a two-dimensional flow: Levy flights and anomalous diffusion. *Physica D* **76**, 70–84 (1994).
17. Panetta, R. L. Zonal jets in wide baroclinically unstable regions: Persistence and scale selection. *J. Atmos. Sci.* **50**, 2073–2106 (1993).
18. Drazin, P. G. & Reid, W. *Hydrodynamic Stability* (Cambridge Univ. Press, Cambridge, 1981).
19. Humphreys, T. D. *Jovian Kármán Vortex Streets as Attractors in Turbulent Zonal Flow*. Thesis, Univ. California, Berkeley (2000).
20. Dowling, T. E. & Ingersoll, A. P. Jupiter's Great Red Spot as a shallow water system. *J. Atmos. Sci.* **46**, 3256–3278 (1989).
21. Marcus, P. S. Vortex dynamics in a shearing zonal flow. *J. Fluid. Mech.* **215**, 393–430 (1990).
22. Pedlosky, J. *Geophysical Fluid Dynamics* (Springer, New York, 1987).
23. Chandrasekhar, S. *Hydrodynamic and Hydromagnetic Stability* (Oxford Univ. Press, Oxford, 1961).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements The author thanks T. Kundu for cloud simulations and X. Asay-Davis, S. Shetty and C.-H. Jiang for calculations of temperature changes. The work was supported by the NASA Origins Program, the NSF Astronomy and Plasma Physics Programs and LANL. Computing resources were supplied by NPACI (supported by the NSF).

Competing interests statement The author declares that he has no competing financial interests.

Correspondence and requests for materials should be addressed to the author (pmarcus@me.berkeley.edu).

The ultimate speed of magnetic switching in granular recording media

I. Tudosa¹, C. Stamm¹, A. B. Kashuba², F. King³, H. C. Siegmann¹, J. Stöhr¹, G. Ju⁴, B. Lu⁴ & D. Weller⁴

¹Stanford Synchrotron Radiation Laboratory, PO Box 20450, Stanford, California 94309, USA

²Landau Institute for Theoretical Physics, Kosygin str. 2, Moscow 117940, Russia

³Stanford Linear Accelerator Center, Stanford University, Stanford, California 94309, USA

⁴Seagate Technology LLC, Pittsburgh, Pennsylvania 15222, USA

In magnetic memory devices, logical bits are recorded by selectively setting the magnetization vector of individual magnetic domains either 'up' or 'down'. In such devices, the fastest and most efficient recording method involves precessional switching^{1–4}: when a magnetic field B_p is applied as a write pulse over a period τ , the magnetization vector precesses about the field until $B_p\tau$ reaches the threshold value at which switching occurs. Increasing the amplitude of the write pulse B_p might therefore substantially shorten the required switching time τ and allow for

faster magnetic recording. Here we use very short pulses of a very high magnetic field⁵ to show that under these extreme conditions, precessional switching in magnetic media supporting high bit densities no longer takes place at well-defined field strengths; instead, switching occurs randomly within a wide range of magnetic fields. We attribute this behaviour to a momentary collapse of the ferromagnetic order of the spins under the load of the short and high-field pulse, thus establishing an ultimate limit to the speed of deterministic switching and magnetic recording.

Our conceptually simple technique, which could also be used to study the dynamics of ferromagnetic spins underlying many applications and promising developments in magnetism^{6–8}, utilizes relativistic electron bunches of energy 28 GeV from the Stanford Linear Accelerator to generate unique short and strong magnetic field pulses^{5,9}. Our magnetic field resembles the field generated by a straight current-carrying wire, with the familiar closed circular magnetic field lines about the beam direction with the field strength decreasing as $1/R$ with the distance R from the centre of the beam. The electron beam is focused to a cross-section of $10.8 \times 7.4 \mu\text{m}$ (full-width at half-maximum) in the x – y plane of the sample surface, perpendicular to the z propagation direction, which lies along the surface normal. Along z , the electron distribution is gaussian with a variance of $\sigma_z = 0.7 \text{ mm}$ in the laboratory frame, giving a pulse duration of $\tau = \sigma_z/c = 2.3 \times 10^{-12} \text{ s}$, where c is the speed with which the electrons travel. For all practical purposes that speed is equal to the speed of light.

With the films magnetized perpendicular to the film plane, the magnetic field B_p and magnetization $\mathbf{M}$ are orthogonal everywhere. This is the optimum geometry to induce a precessional motion of $\mathbf{M}$ about the perpendicular B_p field, which lies in the magnetically hard plane of the film. Once $\mathbf{M}$ has precessed about B_p by an angle large enough to cross the hard plane of the sample, it will continue to relax by itself into the opposite direction. Hence in the end it has switched from one easy direction into the opposite easy direction. If however B_p ceases to exist before $\mathbf{M}$ has reached the hard plane, $\mathbf{M}$ is expected to relax back to its original perpendicular direction, hence no switch is observed. The condition for switching is that the angle of precession $\phi = \omega\tau \geq \pi/2$. As the angular velocity ω is determined by B_p , we obtain the switching condition $B_p\tau \geq \text{const}$.

To test this switching, we prepared 14-nm-thick films of perpendicular granular magnetic recording media of the CoCrPt-type such as recently used in high-density magnetic recording¹⁰. The main condition for high-density recording is that the grains are decoupled so that the medium can sustain narrow transitions between 'up' and 'down' bits. The decoupling of the grains occurs through segregation of Cr to the grain boundaries induced by deposition at elevated temperature. The grain size was determined by X-ray diffraction to be $20 \pm 5 \text{ nm}$. This grain size is so small that the magnetic field B_p is homogeneous over the grain size to better than 0.1% over the R -range of interest. We can then assume that the switching of a grain occurs in a homogeneous applied field. As a substrate, we have used glass with appropriate thin buffer layers, with and without adding a soft magnetic underlayer such as needed in perpendicular recording¹⁰.

The top left panel of Fig. 1 displays contour lines of constant B_p in the film plane. B_p is the peak strength of the gaussian magnetic field pulse calculated from the electron bunch parameters as $B_p = 54.7/R$ where R is measured in micrometres and B_p in tesla. The dark central spot indicates the size of the electron beam focus, close to which no data are obtained owing to beam damage in the sample. The scale is chosen to be the same as in the actual magnetic switching patterns recorded in the other panels of Fig. 1.

Before exposure, the samples were magnetized perpendicular to the film plane into what we shall call the 'up' direction. We recorded patterns on the same sample corresponding to a single shot (pulse),

a pattern corresponding to two shots at the same location, and so on, up to seven consecutive shots per pattern. The time separation of consecutive shots was 1 s. Three weeks after exposure, the perpendicular component of $\mathbf{M}$ was imaged by polar magneto-optic Kerr microscopy. A set of such images is shown in Fig. 1. The spatial resolution of Kerr microscopy is $1\ \mu\text{m}$, so that we integrate over 2,500 grains. The spot in the centre of the pattern is due to beam damage. It extends to roughly twice the beam focus. The increase of the damaged area with the number of shots is due to beam jitter, which is estimated to be $\pm 2\ \mu\text{m}$ per shot only. The grey scale of the images is such that the light regions near the edge of the frames correspond to the initial 'up' state. Darkening indicates that particles in the regions have increasingly switched to the 'down' direction.

It is evident that switching occurs along circular contour lines $B_p = \text{const.}$, with the contrast changing gradually over a distance of tens of micrometres, rather than abruptly. The switching is not reversible, because the second pulse does not return $\mathbf{M}$ to the initial 'up' direction. For odd shot numbers, the dark ring where $\mathbf{M}$ has switched from 'up' to 'down' narrows with increasing number of shots, whereas the outer grey zone corresponding to partially

switched $\mathbf{M}$ expands. With an increasing number of even shots, the central light ring narrows and the outer grey zone expands. This switching behaviour is characteristic of a stochastic process. Starting with a homogeneous magnetic 'up' state, it takes only seven shots to create a random distribution of magnetization directions throughout the large grey zone, where $\mathbf{M}$ is 'up' in some grains but 'down' in others.

To describe the stochastic switching, consider that the angle ϕ of precession may be short of $\pi/2$ needed for switching. The lacking precessional angle may be supplied by random torques or by random initial conditions. We now assume that the probability p of such stochastic events is gaussian, and can be expressed as the probability density of an additional magnetic field Γ :

$$p(\Gamma) = \frac{1}{\Delta B \sqrt{2\pi}} e^{-\frac{\Gamma^2}{2(\Delta B)^2}}$$

The magnetic order parameter $M(B)$ is given by the fraction of particles that switch minus the fraction of particles that do not

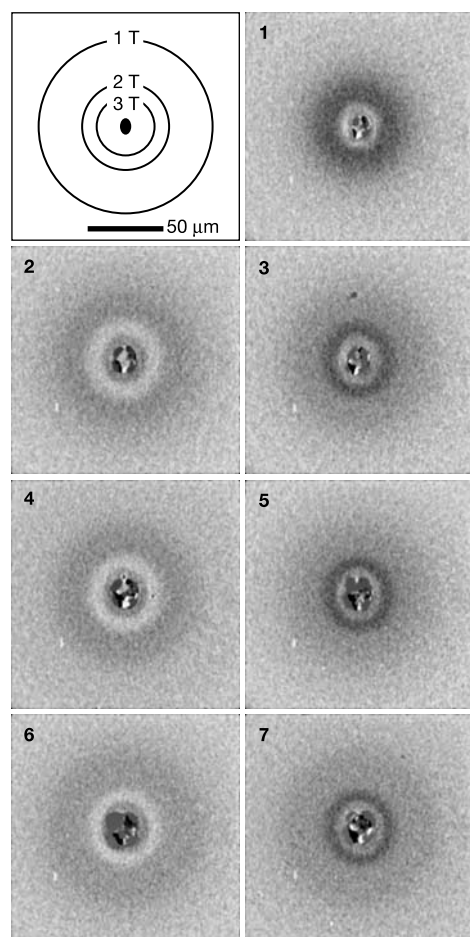


Figure 1 Magneto-optic patterns of magnetization. Diagram at top left, contour lines of constant B_p with area of the electron beam focus in the centre. The numbers on subsequent panels indicate the number of electron bunches (shots) that passed through the sample. Grey contrast is such that the outer light region corresponds to $\mathbf{M}$ in the initial 'up' state. As darkening intensifies, $\mathbf{M}$ has switched increasingly to the 'down' direction. The contrast in the central region at $R < 10\ \mu\text{m}$ is due to beam damage.

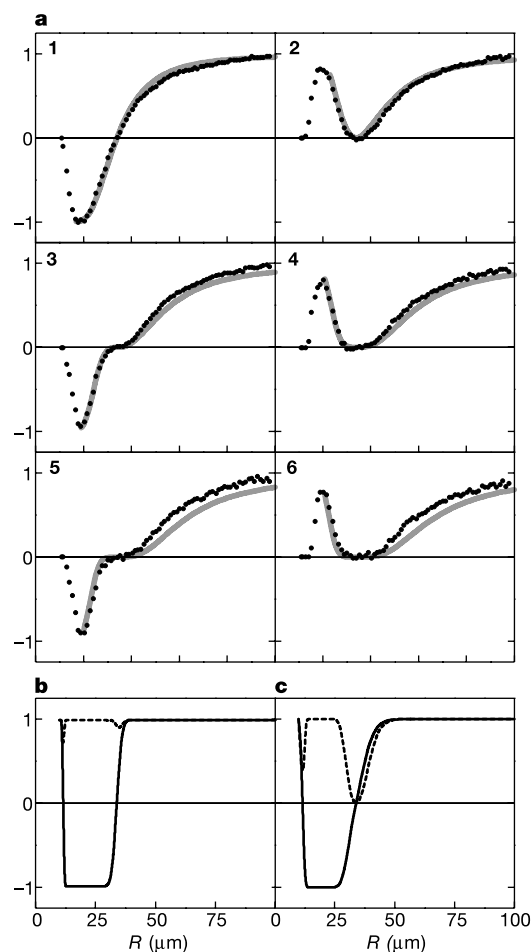


Figure 2 Magnetic order parameter $M_n(R)$. **a**, Data points are from averaged radial cuts of the Kerr images of Fig. 1, setting $M_1(100\ \mu\text{m}) = +1$ and $M_1(20\ \mu\text{m}) = -1$. The grey lines are generated from $M_n(R) = M_1^n(R)$. **b**, Calculated $M_1(R)$ (full line) and $M_2(R)$ (dashed line) with the observed easy-axis dispersion of 5.5° FWHM¹⁰, showing $M_2(R) \equiv M_0 = +1$ in gross contradiction to the experiment. **c**, Calculated $M_1(R)$ (full line) and $M_2(R)$ (dashed line) assuming the excitation of the uniform precession mode (Fig. 3a) corresponding to $K_u V / kT = 40$, where K_u is the uniaxial crystalline anisotropy constant, V the volume of a grain, and kT the Boltzmann factor¹².

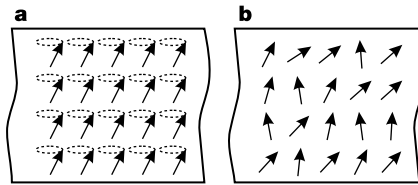


Figure 3 Spin motion in a magnetic grain. **a**, The uniform precession mode of the spins with wavevector $q = 0$. The excitation of this mode determines the long-term stability of the magnetization direction in the grain¹². **b**, A moment in time with non-uniform excitation of the spins. At ambient temperature, these excitations have small amplitude, which however dramatically increases after the field pulse has been applied. Sizeable exchange fields are generated by the angles between neighbouring spins that can account for the random torques operating after the magnetic field pulse.

switch in a single pulse of amplitude B :

$$M(B) = \int_{-B_0-B}^{B_0-B} p(\Gamma) d\Gamma - \int_{B_0-B}^{\infty} p(\Gamma) d\Gamma$$

By choosing the average switching pulse amplitude $B_0 = 1.70$ T and its variance $\Delta B = 0.59$ T we obtain the order parameter $M_1(B_p)$ after the first shot generated from the uniform initial state $M_0 = +1$. Substituting the variable B_p by R yields the radial dependence $M_1(R)$ of the magnetization after the first shot. $M_1(R)$ is plotted for $R > 20 \mu\text{m}$ as a solid line on top of the data points in panel 1 of Fig. 2a. It agrees with the experimental data. The increase of the observed $M_1(R)$ at $R < 20 \mu\text{m}$ indicates the onset of the second switch where M precessed by $\phi \geq 3\pi/2$. We do not analyse this second switch because it occurs close to the beam damage area.

The other distributions of $M_n(R)$ shown as solid lines in panels 2–6 of Fig. 2a are generated by multiplication $M_n(R) = M_{n-1}(R) \times M_1(R) = M_1^n(R)$. This accounts very well for the features observed in the experimental data shown as points in Fig. 2a, despite the fact that raising $M_1(R)$ to the n th power enhances errors. Experimental errors are caused by beam jitter, variations in the number of electrons per bunch, and the uncertainty in the extrapolation of $M_1(R) \rightarrow +1$. At any rate, multiplicative probabilities are the signature of a random variable. Therefore, this analysis reveals that a memory-less process, usually referred to as a Markov uniform stochastic process, dominates the switching.

Different samples have been studied within a range of magnetic parameters such as saturation magnetization and magnetic anisotropy (see Supplementary Information). Information on the switching with older media types such as Co/Pt magnetic multilayers and CoPt alloys is also available^{5,9}. The conclusion emerging from all experiments is that stochastic switching is a general feature of ultrafast precessional magnetization reversal.

To explore causes of the randomness, we carried out various calculations using the Landau–Lifshitz–Gilbert (LLG) equation¹¹. Theoretical results that explore two hypothetical sources of the randomness are shown in Fig. 2b and c. Figure 2b demonstrates that static dispersion of the easy axis of magnetization in the decoupled grains cannot produce anything but deterministic switching reversing M to the original state $M_0 = +1$ in the second shot. This is in gross contradiction to the experiment. Figure 2c explores thermal excitations in terms of the uniform precession mode (illustrated in Fig. 3a). The degree to which the uniform mode is excited is known from the long-term stability of the magnetic bits¹². It induces randomness in the direction of $\mathbf{M}$ before the arrival of the field pulse and indeed generates dispersion of $M(R)$, but the dispersion is much too small to explain the data. The effect of heating of the

sample by the electron pulse can be asserted without calculation. The supersonic heat wave emerging from the point of beam impact requires 10^{-9} s to travel $1 \mu\text{m}$. However, the switching at $R > 20 \mu\text{m}$ is already completed at that time. Similarly, magneto-static coupling between the grains cannot explain the variance of the switching fields because it is small at the end of the field pulse when the spins have precessed close to the hard plane (see Supplementary Information).

However, the thermal fluctuations within a grain also include higher modes in which the spins are not parallel to each other (illustrated in Fig. 3b). At ambient temperature, these fluctuations have small amplitude. Calculations show that the sharply rising field pulse greatly amplifies the pre-existing thermal randomness. The amplification of the thermal fluctuations leads to the observed variance of the switching fields.

The crystalline magnetic anisotropy $H_A = 1.20 \times 10^6 \text{ A m}^{-1}$, the saturation magnetization $M_s = 0.652$ T, and the average switching field $B_0 = 1.7$ T are compatible with the LLG equation if the damping of the magnetization precession in a grain is assumed to be $\alpha = 0.3$. This extremely large damping shows that torque is lost at a high rate to the spin system, proving indeed the excitation of spin fluctuations. It is well known that the spin system is pushed easily into auto-oscillation and chaos as the absorbed power increases^{13,14}. At the end of the field pulse, the non-equilibrium modes illustrated in Fig. 3b exert the postulated random torques. We therefore believe that our experiment reveals ‘fracture of the magnetization’ under the load of the fast and high field pulses, putting an end to deterministic switching as we know it today. □

Received 4 July 2003; accepted 25 February 2004; doi:10.1038/nature02438.

- Gerrits, Th. *et al.* Ultrafast precessional magnetization reversal by picosecond magnetic field pulse shaping. *Nature* **418**, 509–511 (2002).
- Hiebert, W. K., Ballentine, G. E. & Freeman, M. R. Comparison of experimental and numerical micromagnetic dynamics in coherent precessional switching and modal oscillations. *Phys. Rev. B* **65**, 140404(R) (2002).
- Kaka, S. & Russek, S. E. Precessional switching of submicrometer spin valves. *Appl. Phys. Lett.* **86**, 2958–2960 (2002).
- Schumacher, H. W. *et al.* Phase coherent precessional magnetization reversal in microscopic spin valve elements. *Phys. Rev. Lett.* **90**, 017201 (2003).
- Siegmann, H. C. *et al.* Magnetism with picosecond field pulses. *J. Magn. Magn. Mater.* **151**, L8–L12 (1995).
- Rhie, H.-S., Dürr, H. A. & Eberhardt, W. Femtosecond electron and spin dynamics in Ni/W(110) films. *Phys. Rev. Lett.* **90**, 247201 (2003).
- Kiselev, S. I. *et al.* Microwave oscillations of a nanomagnet driven by a spin-polarized current. *Nature* **425**, 380–383 (2003).
- Urazhdin, S. *et al.* Current-driven magnetic excitations in permalloy-based multilayer nanopillars. *Phys. Rev. Lett.* **91**, 146803 (2003).
- Back, C. H. *et al.* Magnetization reversal in ultrashort magnetic field pulses. *Phys. Rev. Lett.* **81**, 3251–3254 (1998).
- Lu, B. *et al.* High anisotropy CoCrPt(B) media for perpendicular magnetic recording. *J. Appl. Phys.* **93**, 6751–6753 (2003).
- Hillebrands, B. & Ounadjela, K. (eds) *Spin Dynamics in Confined Magnetic Structures I* (Springer, Berlin, 2002).
- Weller, D. & Moser, A. Thermal effect limits in ultrahigh-density magnetic recording. *IEEE Trans. Magn.* **35**, 4423–4439 (1999).
- Ebels, U., Buda, L. D., Ounadjela, K. & Wigen, P. E. in *Spin Dynamics in Confined Magnetic Structures I* (eds Hillebrands, B. & Ounadjela, K.) 178–181 (Springer, Berlin, 2002).
- Safonov, V. L. & Bertram, H. N. Spin-wave dynamic reversal in a quasi-single domain magnetic grain. *Phys. Rev. B* **63**, 094419 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank R. Iverson, C. Field and G. J. Collet for their assistance in preparing and carrying out the sample exposure at FFTB, and A. Vaterlaus for help with imaging the magnetic patterns. This work was carried out at the Stanford Linear Accelerator Center, which is supported by the Office of High Energy and Nuclear Physics of the US Department of Energy (DOE). The experimental programme of the SSRL authors is funded by the Office of Basic Energy Sciences of DOE.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.S. (stohr@slac.stanford.edu).

Collapse and rapid resumption of Atlantic meridional circulation linked to deglacial climate changes

J. F. McManus¹, R. Francois², J.-M. Gherardi³, L. D. Keigwin¹ & S. Brown-Leger²

¹Department of Geology and Geophysics, ²Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

³Laboratoire des Sciences du Climat et de l'environnement, Domaine de CNRS, 91198 Gif-sur-Yvette, France

The Atlantic meridional overturning circulation is widely believed to affect climate. Changes in ocean circulation have been inferred from records of the deep water chemical composition derived from sedimentary nutrient proxies¹, but their impact on climate is difficult to assess because such reconstructions provide insufficient constraints on the rate of overturning². Here we report measurements of ²³¹Pa/²³⁰Th, a kinematic proxy for the meridional overturning circulation, in a sediment core from the subtropical North Atlantic Ocean. We find that the meridional overturning was nearly, or completely, eliminated during the coldest deglacial interval in the North Atlantic region, beginning with the catastrophic iceberg discharge Heinrich event H1, 17,500 yr ago, and declined sharply but briefly into the Younger Dryas cold event, about 12,700 yr ago. Following these cold events, the ²³¹Pa/²³⁰Th record indicates that rapid accelerations of the meridional overturning circulation were concurrent with the two strongest regional warming events during deglaciation. These results confirm the significance of variations in the rate of the Atlantic meridional overturning circulation for abrupt climate changes.

²³¹Pa and ²³⁰Th are produced in sea water at constant rates from the radioactive decay of dissolved ²³⁵U and ²³⁴U ($\beta_{\text{Pa}} = (2.45 \pm 0.05) \times 10^{-3} \text{ d.p.m. m}^{-3} \text{ yr}^{-1}$; $\beta_{\text{Th}} = (2.62 \pm 0.05) \times 10^{-2} \text{ d.p.m. m}^{-3} \text{ yr}^{-1}$; $\beta_{\text{Pa}}/\beta_{\text{Th}} = 0.093$; β , production rate; d.p.m., disintegrations per minute). Both are rapidly removed from sea water by adsorption on settling particles, resulting in relatively short residence times in the water column and in the accumulation of excess, or unsupported, ²³⁰Th and ²³¹Pa in the underlying sediments. ²³¹Pa is less rapidly removed than ²³⁰Th and has a residence time (~100–200 yr) approaching the transit time of deep water in the Atlantic basin³. As a result, approximately half of the ²³¹Pa produced in Atlantic water is exported with the North Atlantic Deep Water (NADW) into the Southern Ocean instead of being removed into the Atlantic sediments. On the other hand, ²³⁰Th has a shorter residence time (~20–40 yr), which minimizes its transport from the Atlantic to the Southern Ocean today. Modern North Atlantic sediments thus retain a mean $\text{ex}^{231}\text{Pa}/\text{ex}^{230}\text{Th}_0$ (excess ²³¹Pa and ²³⁰Th activity ratio decay-corrected to the time of deposition; ²³¹Pa/²³⁰Th hereafter) significantly lower than the seawater production ratio of 0.093. For a given scavenging rate, lower rates of meridional overturning circulation (MOC) in the past would result in comparatively less ²³¹Pa export from the Atlantic and in higher sedimentary ²³¹Pa/²³⁰Th, reaching a maximum of 0.093 for a total cessation^{4,5}.

In order to assess changes in MOC during the past 20 kyr, we have analysed a high-accumulation-rate core from Bermuda rise in the deep western subtropical Atlantic (OCE326-GGC5; 33° 42' N, 57° 35' W, 4.55 km). Deglacial stratigraphy was constructed using $\delta^{18}\text{O}$ measurements in individual shells of the planktonic foraminifera *Globorotalia inflata*, and the chronology was established using ¹⁴C dates from monospecific foraminifera samples. A large fraction

of Bermuda rise sediment consists of older fine-grained material laterally transported from the Canadian margin⁶. However, the kinetics of adsorption reactions are believed to be fast enough⁷ that the laterally transported sediment does not retain the ²³¹Pa/²³⁰Th from its initial site of deposition. Instead, ²³¹Pa/²³⁰Th is reset by equilibration with the transporting water, and the ²³¹Pa/²³⁰Th signal at the site of final deposition is synchronous with the deposition of the planktonic foraminifera that were used to establish the chronology. ²³⁰Th and ²³¹Pa were analysed by isotope dilution on a magnetic sector inductively coupled plasma mass spectrometer⁸.

The ²³¹Pa/²³⁰Th ratio in the most recent sediments (~0.055) is consistent with the export from the North Atlantic of ~10% of the ²³⁰Th and ~50% of the ²³¹Pa production by the vigorous modern MOC^{4,5}. Similar values characterize the entire Holocene sections (0.055 ± 0.006 ; 2σ ; $n = 9$), suggesting no major changes in Atlantic MOC during the past 10 kyr (Fig. 1). The slightly lower ratio (0.050) during the early Holocene implies a slightly increased rate of MOC, and a single higher ratio (0.062) in the mid-Holocene (~5 kyr ago) raises the possibility of a brief decline in MOC, but these results must be confirmed with additional high-resolution analyses.

The mean ²³¹Pa/²³⁰Th of sediment deposited during the Last Glacial Maximum (LGM) between 17.5 and 19.8 kyr ago (Fig. 1) is significantly higher (0.068 ± 0.010 ; 2σ ; $n = 12$) than during the Holocene (0.055 ± 0.006 ; 2σ ; $n = 9$). The difference in the mean ²³¹Pa/²³⁰Th between the Holocene and LGM ($\Delta(^{231}\text{Pa}/^{230}\text{Th})_{\text{LGM-Hol}} = 0.013 \pm 0.004$; 95% confidence interval) is at the upper limit of the range of glacial reduction in MOC estimated⁵ (30%) using a broader North Atlantic ²³¹Pa/²³⁰Th database and a circulation–biogeochemistry model. The new results from GGC5 also suggest no more than a ~30–40% slowdown of the MOC in the LGM, and probably less, because shallower overturning and generally higher glacial particle fluxes (Fig. 2) could also contribute to higher sedimentary ²³¹Pa/²³⁰Th.

In contrast, the deglaciation is marked by larger changes in ²³¹Pa/²³⁰Th whose timing suggests that rapid, large-scale oscillations in MOC occurred in concert with regional climatic variations

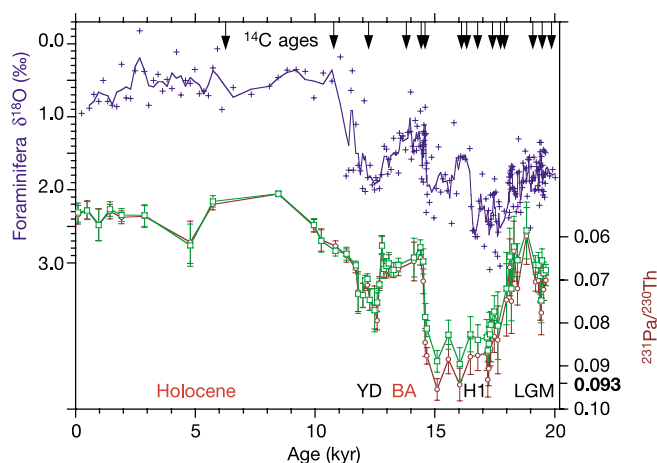


Figure 1 Stable isotope and radiochemical data from sediment core OCE326-GGC5 (33° 42' N, 57° 35' W, 4.55 km). Individual measurements and three-point running mean of planktonic foraminifera $\delta^{18}\text{O}$ (*G. inflata*) plotted with sedimentary ²³¹Pa/²³⁰Th against age (238-based values (red circles) are calculated by subtraction of supported ²³⁰Th and ²³¹Pa estimated from the measured ²³⁸U activity; 232-based values (green squares) are calculated by subtraction of supported ²³⁰Th and ²³¹Pa estimated from the measured ²³²Th activity $\times 0.57$ (see Methods)). Both $\delta^{18}\text{O}$ and ²³¹Pa/²³⁰Th axes are inverted. Arrows indicate ¹⁴C age control points. Labels above lower axis indicate significant climatic intervals: Holocene, warm; YD, Younger Dryas, cooling; BA, Bölling–Allerød, warm; H1, Heinrich event 1, iceberg discharge; LGM, Last Glacial Maximum.

(Fig. 2). The most striking feature of the $^{231}\text{Pa}/^{230}\text{Th}$ profile is the abrupt increase towards the production ratio of 0.093, starting at 17.5 kyr ago and continuing until 15 kyr ago. The data indicate a rapid decline in the export of ^{231}Pa from this site. As the response time of sediment $^{231}\text{Pa}/^{230}\text{Th}$ to changes in circulation is ~ 500 yr (refs 4, 5), the $^{231}\text{Pa}/^{230}\text{Th}$ profile can be interpreted as reflecting a quasi-total and nearly instantaneous cessation of the Atlantic MOC at 17.5 kyr ago. Higher $^{231}\text{Pa}/^{230}\text{Th}$ could also be produced by higher particle flux and/or higher opal concentration in settling particles,

reducing the residence time and fractionation of the two nuclides in the water column⁹. These alternative explanations can be ruled out, however, because clay and carbonate fluxes were lower than during the LGM, and Si/Al indicates that opal content in the sediment is negligible (Fig. 2) (see Methods section). The $^{231}\text{Pa}/^{230}\text{Th}$ signal is thus best interpreted as reflecting a nearly total shutdown of the MOC that lasted more than 2,000 yr.

The onset at 17.5 kyr ago coincides with the timing of Heinrich event H1^{10,11}, a catastrophic iceberg discharge into the North Atlantic, and with additional evidence for freshening at high latitudes¹², and diminished relative influence of northern-source deep waters during that time^{13,14}. Model simulations suggest that the MOC is sensitive to buoyancy forcing through surface salinity perturbations¹⁵, with a sufficiently marked response in some experiments to resemble a 'drop-dead circulation' with negligible overturning¹⁶. It was previously tempting to speculate that the puzzling interval between the onset of the severe climate associated with the H1 iceberg catastrophe (~ 17 kyr ago) and the eventual dramatic warming¹⁷ of the Bølling–Allerød (~ 15 kyr ago) might be the result of such a drop-dead circulation¹⁸. The $^{231}\text{Pa}/^{230}\text{Th}$ results from GGC5 provide the best indication yet that such a scenario occurred during that time.

At the same time as the rise in $^{231}\text{Pa}/^{230}\text{Th}$, there is also an increase in planktonic foraminifera $\delta^{18}\text{O}$ (Fig. 2). The rapidity of the $\delta^{18}\text{O}$ shift and the fact that it occurs subsequent to the timing of the glacial maximum¹⁹ rule out the possibility that it reflects an increase in global ice volume. It must therefore represent a near-surface

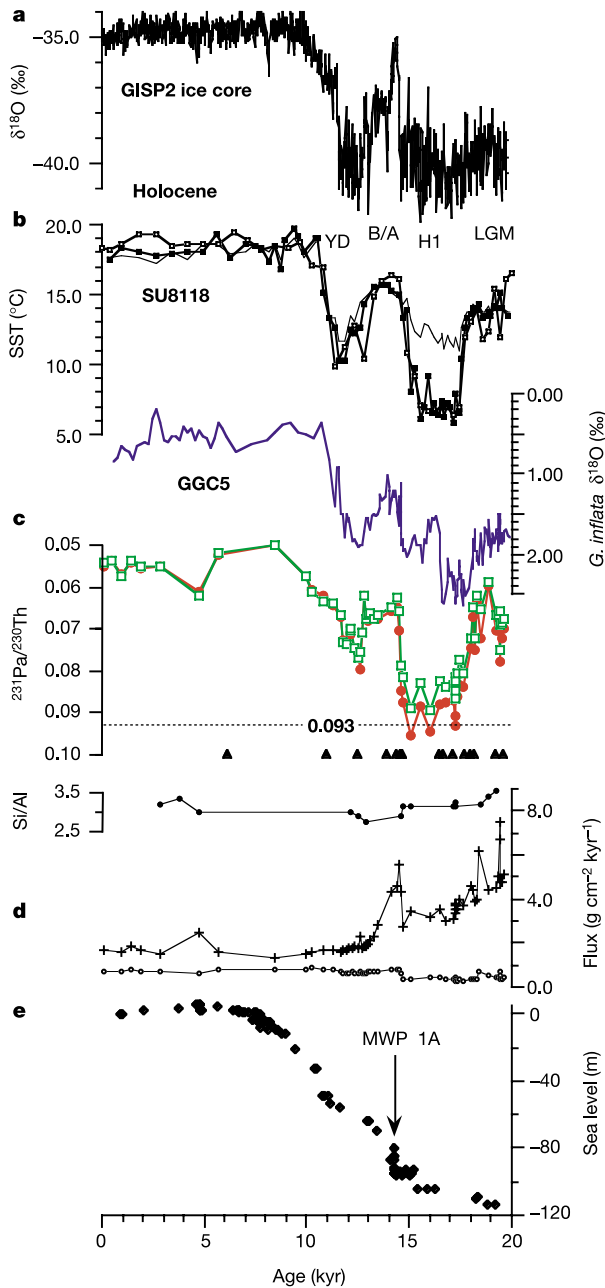


Figure 2 Comparison of sedimentary data from the subtropical Atlantic with ice-core, subpolar deep-sea core, and coral records for the past 20 kyr. **a**, Greenland $\delta^{18}\text{O}$ record²⁰. **b**, SST estimates from the subpolar North Atlantic based on planktonic foraminiferal assemblage²⁹ (open symbols) and two calibrations for alkenone unsaturation ratios¹¹. **c**, Three-point running mean of planktonic foraminifera $\delta^{18}\text{O}$ plotted with sedimentary $^{231}\text{Pa}/^{230}\text{Th}$ against calendar age in core GGC5. Black triangles indicate ^{14}C age control points. **d**, ^{230}Th -normalized fluxes of carbonate and terrigenous sediment. **e**, Sea-level rise due to melting ice sheets^{21,22}.

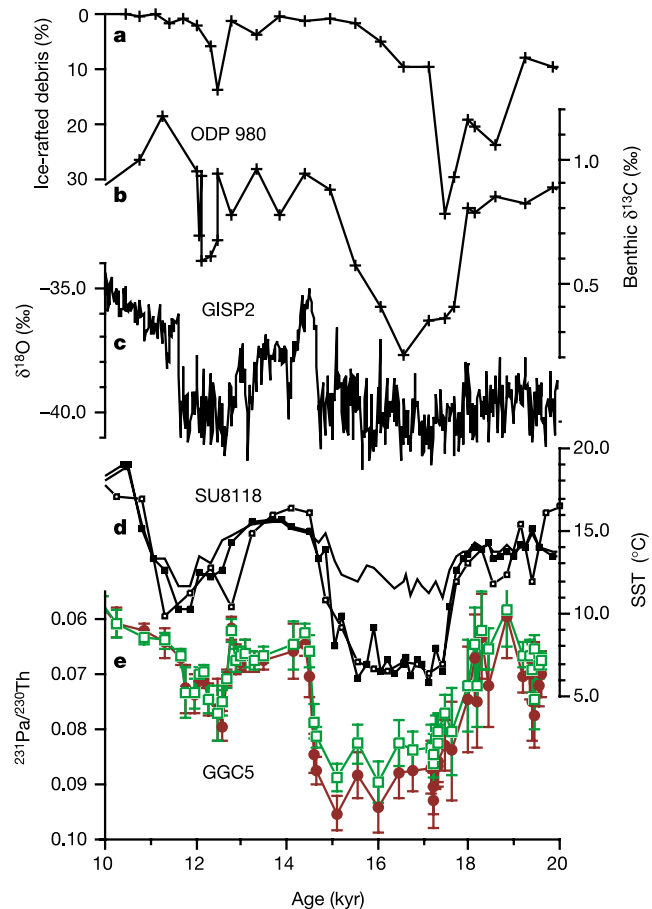


Figure 3 Deglacial records of ocean circulation and climate in the North Atlantic region. **a**, Ice-rafted debris record from ODP Site 980²⁵. **b**, Benthic foraminifera $\delta^{13}\text{C}$ from ODP Site 980²⁵. **c**, $\delta^{18}\text{O}$ in Greenland ice²⁰. **d**, SST estimates from the subpolar North Atlantic^{11,29}. **e**, Sedimentary $^{231}\text{Pa}/^{230}\text{Th}$ from GGC5.

cooling and/or increase in local salinity at the core location. The amplitude of the $\delta^{18}\text{O}$ increase (0.75–1.0‰) implies that sea surface temperature (SST) at this subtropical site may have been as much as 3–4 °C lower at the time of H1 than during the LGM. Minimum subpolar SST¹¹ and colder air temperatures over Greenland²⁰ are also evident at this time (Fig. 2), underscoring the fact that the LGM should not be viewed simply as a climatic extreme everywhere. The synchrony between changes in $^{231}\text{Pa}/^{230}\text{Th}$ and foraminifera $\delta^{18}\text{O}$ suggests a connection between the Atlantic MOC and local surface hydrography, and the inferred near-cessation of MOC during early deglaciation appears to be directly linked to the freshening and increased buoyancy of Northern Atlantic surface water^{10–12}. These findings thus support earlier suggestions that melt water associated with catastrophic iceberg discharge freshened the high-latitude surface ocean, stabilized the water column, and weakened the Atlantic MOC^{13,14,18}. Our results indicate that the effect was dramatic, resulting in a near-total collapse of the Atlantic MOC and substantial regional cooling.

The period of high $^{231}\text{Pa}/^{230}\text{Th}$ was terminated by a decrease to less than 0.065, suggesting an abrupt resumption of the Atlantic MOC at a time (~14.7 kyr ago) corresponding to the onset of the Bølling–Allerød warm period (Figs 1, 2). The timing of this resumption is shortly before the most rapid deglacial reduction of continental ice and sea-level rise, meltwater pulse 1a (MWP 1A), which has been dated at ~14.0 kyr ago^{21,22} (Figs 2, 3). Reinvigoration of the oceanic heat transport to high northern latitudes may have combined with increasing insolation to accelerate the melting of the Laurentide ice sheet, resulting in MWP 1A. Previous evidence has pointed to a large, rapid, high-latitude warming at this time^{17,23}. Our new results provide support for the idea that the rapidity and magnitude of this warming resulted in large part from the reinvigoration of the Atlantic MOC.

After reaching a minimum at 0.065 at 14.2 kyr ago, $^{231}\text{Pa}/^{230}\text{Th}$ remained nearly constant (0.065 ± 0.004 ; 2σ ; $n = 8$) for ~1,500 yr (14.2–12.5 kyr ago), when the temperature recorded in Greenland ice decreased in several steps (Fig. 3). This stepwise cooling, starting at ~14 kyr ago, does not seem to be the result of a decrease in the Atlantic MOC.

Melt water associated with H1 appears to have been particularly efficient at curtailing meridional overturning, most probably owing to the delivery of drifting icebergs closer to the sensitive locations of glacial deep-water formation. No equivalent suppression of MOC accompanied MWP 1A, even though this event may have involved much larger volumes of fresh water. One possible explanation is that this melt water originated primarily from Antarctica with only a modest contribution from the Northern Hemisphere²⁴. Another possibility is that runoff from the Laurentide ice sheet underwent substantial mixing with sea water before reaching the zones of deep convection. The site of deep-water production may also have migrated north of the Greenland–Iceland–Scotland ridge, farther from the influence of melt water from North America.

At 12.7 kyr ago, $^{231}\text{Pa}/^{230}\text{Th}$ increased abruptly, although not to the same level as during H1 (Fig. 2), but implying at least a partial reduction of the Atlantic MOC. This shift coincides with the beginning of the Younger Dryas, and is associated with an increase in the planktonic $\delta^{18}\text{O}$, probably reflecting a decrease in subtropical SST (Fig. 2). MWP 1A occurred too early to be a direct cause of a circulation change or cooling during the Younger Dryas. By contrast, evidence from the subpolar North Atlantic²⁵ points to a link between enhanced deposition of ice-rafted debris and a decline in $\delta^{13}\text{C}$ at this time (Fig. 3). The Younger Dryas may therefore have been triggered by iceberg discharge, in a manner similar to H1.

Unlike the earlier near-shutdown, however, the decline in MOC at 12.7 kyr ago did not last thousands of years. The rapid increase in $^{231}\text{Pa}/^{230}\text{Th}$ was immediately followed by decreases in both $\delta^{18}\text{O}$ and $^{231}\text{Pa}/^{230}\text{Th}$. This pattern may indicate a gradual re-acceleration of the Atlantic MOC, as suggested by the $\Delta^{14}\text{C}_{\text{atm}}$ reconstructed

from the Cariaco basin²⁶. In general, the $^{231}\text{Pa}/^{230}\text{Th}$ results are in agreement with $\Delta^{14}\text{C}_{\text{atm}}$ interpretations of deglacial changes in MOC, with a sharp decline during H1, a reinvigoration during the Bølling–Allerød to very similar values as the LGM, and then brief, sharp decline and gradual increase within the Younger Dryas and continuing into the Holocene^{26–28}.

At face value, the $^{231}\text{Pa}/^{230}\text{Th}$ suggests only a partial weakening of MOC during the Younger Dryas, in contrast to the near shutdown during H1. In this case, the comparison with estimated changes in SST^{11,29} suggests strong coupling between the Atlantic MOC and climate in at least part of the North Atlantic region for the past 20 kyr (Fig. 2). However, because of the brevity of the slowdown and the response time of $^{231}\text{Pa}/^{230}\text{Th}$ to circulation shutdowns (~500 yr), the Younger Dryas reduction in the rate of MOC may have been larger than indicated by the maximum in $^{231}\text{Pa}/^{230}\text{Th}$, and the possibility of a total, albeit very brief, shutdown cannot be ruled out.

The subsequent reinvigoration of the Atlantic MOC was more gradual than during the Bølling warming (Fig. 2), although lower sedimentation rates in this interval of GGC5 may have resulted in the smoothing of a more abrupt transition by bioturbation. Thereafter, the acceleration of the Atlantic MOC continued uninterrupted to its modern strength, which has largely prevailed throughout the Holocene. The deglacial transition thus contains the most dramatic changes in the rate of Atlantic MOC as well as the largest climatic oscillations of the past 20 kyr. Marked declines in MOC were associated with both the H1 and Younger Dryas millennial cooling events, and in each case the rapid subsequent increase in MOC was accompanied by abrupt warming. A remaining question is why the LGM reduction in MOC was limited to 30–40% or less, while briefer events were substantially larger. One possibility is that the steady-state rate of the overturning circulation may be fixed by the energy available for diapycnal mixing³⁰, whereas transient salinity perturbations at the surface may have a shorter-term but potentially more dramatic impact on MOC and its climatically significant heat transport. An improved understanding of the role of such transient perturbations should thus remain a high priority for future investigations. □

Methods

Uranium, thorium and protactinium were measured by isotope dilution on a Finnigan MAT Element 1 single collector, sector field, inductively coupled plasma mass spectrometer⁸. The samples were ground with mortar and pestle, then spiked with ^{233}Pa and ^{229}Th before total dissolution in HNO_3 , HF and HClO_4 . An aliquot of the resulting solution was removed, spiked with ^{236}U and ^{229}Th , and analysed for ^{238}U and ^{232}Th . The remaining solution was used for ^{231}Pa and ^{230}Th separation by ion-exchange chromatography. The internal precision on the measured $^{231}\text{Pa}/^{230}\text{Th}$ was usually better than 1.5% (2σ level). Reproducibility for full replicate analyses (dissolution, chromatographic separation and spectrometry) was usually better than 5% (2σ level) on the measured $^{231}\text{Pa}/^{230}\text{Th}$. Excess activities were obtained using two independent corrections for the supported detrital portion of the total ^{230}Th and ^{231}Pa measured. One estimate is based on the total ^{238}U activity and the other is based on the total ^{232}Th activity, using a $^{238}\text{U}/^{232}\text{Th}$ activity ratio (0.57) within a range typical for Atlantic sediment (0.6 ± 0.2)³⁵. Each estimate was corrected for radioactive decay of the unsupported nuclides since the time of sediment deposition determined from the ^{14}C -based age-model. The ^{238}U correction is sensitive to the deposition of authigenic U, and the ^{232}Th correction may be sensitive to variations in sediment source or type. The good agreement of the two estimates throughout the record (Fig. 1) indicates that these potentially confounding factors are not important here. Thus, no calculation of post-depositional ingrowth of ^{230}Th or ^{231}Pa from authigenic uranium was required.

Sedimentary fluxes

Mass fluxes were calculated from the decay-corrected activity of excess ^{230}Th per gram of sediment and the expected ^{230}Th flux to the sea floor, given the concentration of ^{234}U in sea water and the water depth³². Fluxes of CaCO_3 and terrigenous sediment of all sizes were calculated based on the bulk mass flux and the percentage of CaCO_3 , measured by coulometer. Uncertainties in the mass flux estimates averaged <6%. Two other sediment types, authigenic precipitates and opal, occurred in negligible amounts, although the potential influence of opal on $^{231}\text{Pa}/^{230}\text{Th}$ led us to explore several means of assessing its presence. Bulk Si/Al (Fig. 2) measured on a representative sample set (3.1 ± 0.2) was found to be slightly below that of the upper continental crust (3.8) and similar to the average post-Archaeozoic Australian terrigenous shale (2.9). Fine-fraction aliquots were prepared for micropalaeontological analysis of diatoms and examined with a compound

light microscope, but too few frustules occurred to allow quantification. The coarse, sand-sized fraction of the samples was also examined visually. In particular, the sediments greater than 150 μm , representing less than 1% of the sample weight, were examined for large diatoms in the deglacial interval of the core. A portion of these samples contained the diatom *Ethmodiscus rex*, although their occurrence does not covary with $^{231}\text{Pa}/^{230}\text{Th}$, and none were observed from 190–220 cm, when $^{231}\text{Pa}/^{230}\text{Th}$ reached maximum values. Although it cannot be entirely ruled out that additional diatom frustules or other opal may have been deposited originally and subsequently dissolved, the described analyses indicate that opal concentration in the sediments is negligible and cannot account for the $^{231}\text{Pa}/^{230}\text{Th}$ signal.

Stratigraphy and chronology

The stratigraphy for the core was constructed using $\delta^{18}\text{O}$ measurements on single shells of the planktonic foraminifera *G. inflata*. Samples of 0.5–1.0 cm thickness were taken from the core and dried. The dry samples were weighed, immersed in distilled water, and wet-sieved using a 63- μm mesh. The >63- μm fraction was dried again, and dry-sieved using a 150- μm mesh. Individual specimens of *G. inflata* were hand-selected from the >150- μm fraction using 25–50 $\times$ magnification under a binocular microscope. Each specimen was acidified and analysed on a partially automated VG 903 stable isotope mass spectrometer. The resulting oxygen isotope ratios were calibrated to $\delta^{18}\text{O}$ using NBS-19 and are reported relative to the Pee Dee Belemnite (PDB) standard. A chronology for the core was established using fourteen ^{14}C dates from monospecific foraminifera samples (typically *G. inflata*). An abundance maximum of dextral coiling *Neogloboquadrina pachyderma* was used to obtain a date for the Younger Dryas interval, and in one case *Globigerinoides ruber* was measured for ^{14}C along with *G. inflata*. The age difference between the two species is less than the measurement uncertainty. A reservoir correction of 400 yr was applied to the measured ^{14}C , and the ages were converted to calendar years using the CALIB 4.3 calibration program³³.

Received 2 January; accepted 16 March 2004; doi:10.1038/nature02494.

- Boyle, E. A. & Keigwin, L. North Atlantic thermohaline circulation during the past 20,000 years linked to high-latitude surface temperature. *Nature* **330**, 35–40 (1987).
- Legrand, P. & Wunsch, C. Constraints from paleotracer data on the North Atlantic circulation during the last glacial maximum. *Paleoceanography* **10**, 1011–1045 (1995).
- Broecker, W. S. A revised estimate of the radiocarbon age of the North Atlantic Deep Water. *J. Geophys. Res.* **84**, 3218–3226 (1979).
- Yu, E.-F., Francois, R. & Bacon, M. P. Similar rates of modern and last-glacial ocean thermohaline circulation inferred from radiochemical data. *Nature* **379**, 689–694 (1996).
- Marchal, O., Francois, R., Stocker, T. F. & Joos, F. Ocean thermohaline circulation and sedimentary $^{231}\text{Pa}/^{230}\text{Th}$ ratio. *Paleoceanography* **15**, 625–641 (2000).
- Ohkouchi, N., Eglinton, T. L., Keigwin, L. D. & Hayes, J. M. Spatial and temporal offsets between proxy records in a sediment drift. *Science* **298**, 1224–1227 (2002).
- Bacon, M. P. & Anderson, R. F. Distribution of thorium isotopes between dissolved and particulate forms in the deep sea. *J. Geophys. Res.* **87**, 2045–2056 (1982).
- Choi, M.-S. *et al.* Rapid determination of ^{230}Th and ^{231}Pa in seawater by desolvated-micronebulization inductively-coupled magnetic sector mass spectrometry. *Mar. Chem.* **76**, 99–112 (2001).
- Chase, Z., Anderson, R. F., Fleisher, M. Q. & Kubik, P. W. The influence of particle composition and particle flux on scavenging of Th, Pa and Be in the ocean. *Earth Planet. Sci. Lett.* **204**, 215–229 (2002).
- Bond, G. *et al.* Evidence for massive discharge of icebergs into the glacial Northern Atlantic. *Nature* **360**, 245–249 (1992).
- Bard, E., Rostek, F., Turo, J.-L. & Gendreau, S. Hydrological impact of Heinrich Events in the Subtropical Northeast Atlantic. *Science* **289**, 1321–1323 (2000).
- Duplessy, J.-C. *et al.* Changes in surface salinity of the North Atlantic Ocean during the last deglaciation. *Nature* **358**, 485–488 (1993).
- Vidal, L. *et al.* Evidence for changes in the North Atlantic Deep Water linked to meltwater surges during the Heinrich events. *Earth Planet. Sci. Lett.* **146**, 13–27 (1997).
- Eliot, M., Labeyrie, L. & Duplessy, J.-C. Changes in North Atlantic deep-water formation associated with the Dansgaard-Oeschger temperature oscillations (60–10 ka). *Quat. Sci. Rev.* **21**, 1153–1165 (2002).
- Rahmstorf, S. Bifurcations of the Atlantic thermohaline circulation in response to changes in the hydrological cycle. *Nature* **378**, 145–149 (1995).
- Manabe, S. & Stouffer, R. J. Two stable equilibria of a coupled ocean-atmosphere model. *J. Clim.* **1**, 841–866 (1988).
- Severinghaus, J. P. & Brook, E. J. Abrupt climate change at the end of the last glacial period inferred from trapped air in polar ice. *Science* **286**, 930–934 (1999).
- Broecker, W. S. Massive iceberg discharges as triggers for global climate change. *Nature* **372**, 421–424 (1994).
- Mix, A. C., Bard, E. & Schneider, R. Environmental processes of the ice age: Land, oceans, glaciers (EPIC). *Quat. Sci. Rev.* **20**, 627–657 (2001).
- Groote, P. M., Stuiver, M., White, J. W. C., Johnsen, S. & Jouzel, J. Comparison of oxygen isotope records from the GISP2 and GRIP Greenland ice cores. *Nature* **366**, 552–554 (1993).
- Fairbanks, R. G. A 17,000-year glacio-eustatic sea level record: influence of glacial melting rates on the Younger Dryas event and deep-ocean circulation. *Nature* **342**, 637–642 (1989).
- Bard, E. *et al.* Deglacial sea-level record from Tahiti corals and the timing of global meltwater discharge. *Nature* **382**, 241–244 (1996).
- Lehman, S. J. & Keigwin, L. D. Sudden changes in North Atlantic circulation during the last deglaciation. *Nature* **356**, 757–762 (1992).
- Clark, P. U., Mitrovic, J. X., Milne, G. A. & Tamiseia, M. E. Sea-level fingerprinting as a direct test for the source of global meltwater pulse 1A. *Science* **295**, 2438–2441 (2002).
- McManus, J. F., Oppo, D. W. & Cullen, J. L. A 0.5 million year record of millennial-scale climate variability in the North Atlantic. *Science* **283**, 971–975 (1999).
- Hughen, K. A. *et al.* Deglacial changes in ocean circulation from an extended radiocarbon calibration. *Nature* **391**, 65–68 (1998).

- Clark, P. U., Pisias, N. G., Stocker, T. & Weaver, A. The role of the thermohaline circulation in abrupt climate change. *Nature* **415**, 863–869 (2002).
- Hughen, K. A. *et al.* ^{14}C Activity and global carbon cycle changes over the past 50,000 years. *Science* **303**, 202–207 (2004).
- Waelbroeck, C. *et al.* Improving past sea surface temperature estimates based on planktonic fossil faunas. *Paleoceanography* **13**, 272–283 (1998).
- Munk, W. & Wunsch, C. Abyssal recipes II: energetics of tidal and wind mixing. *Deep-Sea Res.* **1**, 45, 1976–2009 (1998).
- McManus, J. F., Anderson, R. F., Broecker, W. S., Fleisher, M. Q. & Higgins, S. M. Radiometrically determined fluxes in the sub-polar North Atlantic during the last 140,000 years. *Earth Planet. Sci. Lett.* **135**, 29–43 (1998).
- Francois, R., Frank, M., Rutgers van der Loeff, M. M. & Bacon, M. P. ^{230}Th -normalization: An essential tool for interpreting sedimentary fluxes during the late Quaternary. *Paleoceanography* **19**, doi:10.1029/2003PA000939 (2004).
- Stuiver, M. *et al.* INTCAL98 radiocarbon age calibration. *Radiocarbon* **40**, 1041–1083 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This study was improved by assistance and input from O. Marchal, M. Bacon, W. Curry, L. Labeyrie, R. Anderson, N. Ohkouchi and T. Eglinton. Technical expertise was provided by A. Fleer, E. Roosen, L. Zou, S. Manganini, E. Frank, L. Ball, D. Schneider, I. Grigorov, S. Benetti, M. Jeglinski and A. Edwards. Support for this research was provided in part by the US-NSF OCE and INT programmes, the France CNRS, WHOI-OCCI and Mellon awards, and the Comer Science and Education Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.F.M. (jmcmanus@whoi.edu).

Dislocation creep in MgSiO₃ perovskite at conditions of the Earth's uppermost lower mantle

Patrick Cordier^{1,2}, Tamás Ungár³, Lehel Zsoldos³ & Géza Tichy⁴

¹Laboratoire de Structure et Propriétés de l'Etat Solide, UMR CNRS 8008, Université des Sciences et Technologies de Lille, F-59655 Villeneuve d'Ascq Cedex, France

²Bayerisches Geoinstitut, Universität Bayreuth, D-95440 Bayreuth, Germany

³Department of General Physics,

⁴Department of Solid State Physics, Eötvös University Budapest, H-1518 POB 32, Budapest, Hungary

Seismic anisotropy provides an important observational constraint on flow in the Earth's deep interior. The quantitative interpretation of anisotropy, however, requires knowledge of the slip geometry of the constitutive minerals that are responsible for producing rock fabrics. The Earth's lower mantle is mostly composed of (Mg, Fe)SiO₃ perovskite¹, but as MgSiO₃ perovskite is not stable at high temperature under ambient pressure, it has not been possible to investigate its mechanical behaviour with conventional laboratory deformation experiments. To overcome this limitation, several attempts were made to infer the mechanical properties of MgSiO₃ perovskite on the basis of analogue materials^{2–7}. But perovskites do not constitute an analogue series for plastic deformation, and therefore the direct investigation of MgSiO₃ perovskite is necessary. Here we have taken advantage of recent advances in experimental high-pressure rheology⁸ to perform deformation experiments on coarse-grained MgSiO₃ polycrystals under pressure and temperature conditions of the uppermost lower mantle. We show that X-ray peak broadening measurements developed in metallurgy can be adapted to low-symmetry minerals to identify the elementary deformation mechanisms activated under these conditions. We conclude that, under uppermost lower-mantle conditions, MgSiO₃ perovskite

deforms by dislocation creep and may therefore contribute to producing seismic anisotropy in rocks at such depths.

(Mg, Fe)SiO₃ perovskite consists of a corner-linked network of SiO₆ octahedra with Mg atoms in a cavity formed by eight octahedra. The structure is orthorhombic (space group *Pbnm*; ref. 9) under mantle conditions. This structure differs slightly from the ideal cubic perovskite (space group *Pm3m*) in that the octahedra are tilted and the Mg atoms are displaced from the centres of their octahedral sites (Fig. 1). The pseudo-cubic lattice is often used to describe the defects, as it provides a unified frame for all perovskites. In the following, the pseudo-cubic setting will be referred to by the subscript 'c', while subscript 'o' will be used to refer to the actual orthorhombic symmetry. It appears from deformation studies on analogue compounds that perovskites can deform with dislocations having $\langle 100 \rangle_c$ or $\langle 110 \rangle_c$ Burgers vectors. Slip system $\langle 1\bar{1}0 \rangle_c \{110\}_c$ seems to be activated from room temperature and is controlled by glide, whereas $\langle 100 \rangle_c \{001\}_c$ slip appears to be characteristic of high temperature only and is controlled by climb.

In this study, deformation experiments were performed under uppermost lower-mantle conditions on polycrystals of silicate perovskite with the enstatite MgSiO₃ composition. As starting material, we used a synthetic MgSiO₃ glass prepared from high-purity oxides. Perovskite synthesis was achieved in a multianvil apparatus at 22–23 GPa, 1,600–2,000 °C for 4–6 h. Visual examination of recovered perovskite (as checked by Raman spectroscopy) specimens revealed large (up to several hundred micrometres)

transparent colourless grains. The specimen was then placed in a second 10-mm octahedral cell assembly modified to induce plastic deformation¹⁰ in compression at high pressure and high temperature. The experimental conditions of the deformation experiments are 25 GPa and 1,400 °C for 1 h. After deformation, the specimen was recovered, and polished thin sections were prepared. In the optical microscope, deformation is shown in the form of dense deformation bands that shear the pre-existing twins (Fig. 2).

Beyond the achievement of plastic deformation experiments under extreme pressures, the challenge with silicate perovskite is to gain information on the deformation mechanisms activated. Indeed, transmission electron microscopy (which is the best suited technique for crystal defect analysis) can hardly be used with silicate perovskite (especially with defective crystals) because of the high sensitivity of this material to electron irradiation. To overcome this limitation, we have characterized dislocations' Burgers vectors through the analysis of the fine structure of X-ray diffraction peak broadening. Peak broadening due to lattice distortions is anisotropic as a function of the *hkl* indices¹¹. As in electron microscopy, where the contrast of dislocations is strong or disappears when the

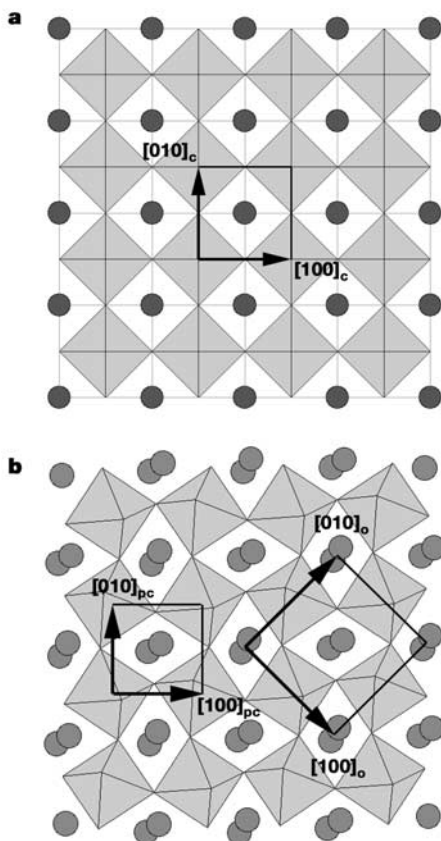


Figure 1 Crystal structure of silicate perovskite. **a**, Ideal (cubic) perovskite structure viewed along $[001]$ with a unit cell representation. **b**, Structure of orthorhombic MgSiO₃ perovskite viewed along the longer axis: $[001]_o$ (the subscript 'o' stands for orthorhombic). An orthorhombic unit cell is represented on the right side. The structure can be approximately described within a pseudo-cubic unit cell (shown on the left side and referenced by the subscript 'pc') with $a_{pc} \approx a_o/\sqrt{2} \approx b_o/\sqrt{2} \approx c_o/2$.

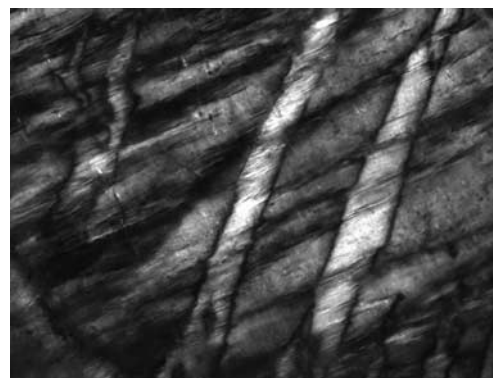


Figure 2 Microstructure. Optical micrograph of MgSiO₃ perovskite deformed at 25 GPa and 1,400 °C in the multianvil apparatus. The twins are sheared by dense deformation bands. The micrograph is 150 μm across.

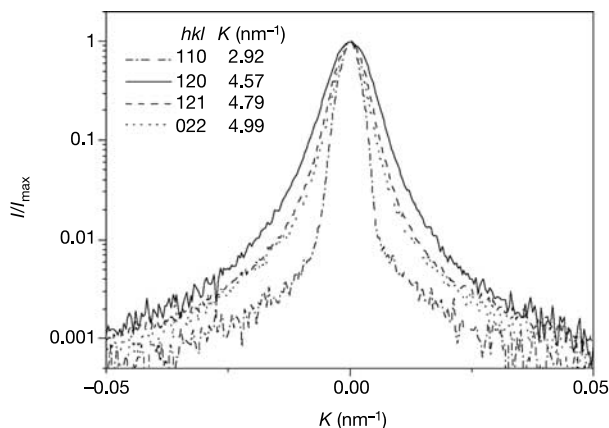


Figure 3 X-ray peak broadening. Profiles of 110, 120, 121 and 022 X-ray diffraction peaks, plotted with logarithmic intensity scales normalized to the maximum intensity versus the reciprocal space coordinate $K (= (2\cos\theta_B/\lambda)\Delta\theta$, where θ_B is the Bragg angle, λ is the wavelength of the X-rays and $\Delta\theta$ is the diffraction angle measured from θ_B). The profiles were measured by a special high-resolution double crystal diffractometer with negligible instrumental effects, operated at the Cu fine focus of a Nonius rotating anode. The parallel beam produced by a symmetrically cut Ge monochromator had a footprint of about 0.1×0.6 mm² on the specimen.

Table 1 Possible dislocations

No.	Slip system	Modulus (Å)	Line direction
1	[100](010)	4.77	[100] screw
2	[100](010)	4.77	[001] edge
3	[100](001)	4.77	[100] screw
4	[100](001)	4.77	[010] edge
5	[010](100)	4.93	[010] screw
6	[010](100)	4.93	[001] edge
7	[010](001)	4.93	[010] screw
8	[010](001)	4.93	[100] edge
9	[001](100)	6.89	[001] screw
10	[001](100)	6.89	[010] edge
11	[001](010)	6.89	[001] screw
12	[001](010)	6.89	[100] edge
13	110	~6.86	[110] screw
14	110	~6.86	[001] edge
15	[110](001)	~6.86	[110] screw
16	[110](001)	~6.86	[110] edge
17	110	~6.86	[110] screw
18	110	~6.86	[001] edge
19	[110](001)	~6.86	[110] screw
20	[110](001)	~6.86	[110] edge

Dislocations and slip systems considered in the present X-ray peak broadening analysis. All of them are considered and labelled in the orthorhombic framework. The lengths (moduli) of the Burgers vectors are also indicated.

Burgers and diffraction vectors are parallel or perpendicular, X-ray diffraction peaks become broader or narrower. This anisotropy of broadening can be summarized in a dislocation contrast factor, C , which depends on the relative orientation between the Burgers and line vectors of dislocations and the diffraction vector, and on the elastic constants of the crystal^{12–16}. Dislocation contrast factors can be calculated numerically¹⁷. The technique originally developed for cubic and hexagonal metals is here developed for minerals with an orthorhombic symmetry. The whole profile fitting procedure, using the hkl dependence of peak broadening, provides (1) experimental values of the dislocation contrast factors, C_m , (2) the dislocation density and character (edge or screw) and (3) the size distribution of crystallites¹⁸. In a plastically deformed single crystal, a large number of C_m values corresponding to different crystal orientations can be measured. Matching these with the numerically calculated theoretical dislocation contrast factors, C_{th} , can provide information about the dislocation structure in terms of Burgers vector types and slip systems, shedding light on the plastic behaviour of the crystal itself—in the present case, $MgSiO_3$ perovskite.

In the deformed specimen, a large single grain of diameter about 0.5 mm has been identified and investigated as a single crystal. The profiles of 11 single-crystal reflections (110, 111, 112, 021, 022, 023, 121, 122, 123, 211 and 006) have been recorded. Four typical diffraction profiles are shown in Fig. 3; the orthorhombic indices are listed in the sequence of increasing reciprocal space coordinate value $K = 2\sin\theta_B/\lambda$, where θ_B is the Bragg angle and λ is the X-ray

Table 3 Identified dislocations

Dislocation type	b	l	n	<i>f</i>
Screw	010	010	–	0.43 ± 0.04
Edge	100	010	001	0.37 ± 0.04
	010	100	001	0.20 ± 0.02

Burgers and line vectors, **b** and **l**, of the dislocations, and normal vectors of the slip planes, **n**, considered in the contrast factor matching, and volume fractions, *f*, provided by the weighted least squares fitting.

wavelength. The first peak (110) is the narrowest, the second (120) is the broadest, and the last one is narrower than the third one. The sequence of broadening differs strongly from the sequence of diffraction, indicating a strong strain anisotropy. Using the whole profile fitting procedure¹⁸, experimental dislocation contrast factors, C_m , have been determined from each Bragg reflection. They have been compared to the theoretical contrast factors for potentially important dislocations in perovskite, listed in Table 1. A detailed investigation of strain anisotropy enabled us to reduce the 20 possible dislocations to 8. For example, those dislocations that have large contrast factors corresponding to one of the narrowest 006 reflections were excluded from further analysis. The volume fraction of these eight possible dislocations, f_j , is determined by the weighted least squares procedure:

$$\sum_{j=1}^{11} \left(C_m - \sum_{j=1}^8 f_j C_{calc}^j \right)^2 = \text{minimum} \quad (1)$$

where $i = 1$ to 11 is for the eleven Bragg reflections. The result of the weighted least squares fitting is listed in Tables 2 and 3. Table 2 gives the measured and calculated dislocation contrast factors, C_m and C_{calc} , respectively, and Table 3 shows the only dislocations that obtained non-vanishing volume fractions during the analysis according to equation (1). The C_{calc} values in Table 2 correspond to the dislocations in Table 3. The fitting procedure was unambiguous. Table 3 shows that all the activated dislocations have Burgers vectors of **b** = $\langle 100 \rangle_o$ type. The $\langle 110 \rangle_o$ type dislocations, on the other hand, are not activated.

The rheological data on various analogue materials with perovskite structures can be compared when crystallography is expressed in the pseudo-cubic frame. It is also necessary to normalize temperatures with melting temperatures and stress by the shear modulus. After proper normalization, the creep laws for the $\langle 110 \rangle_c$ slip systems for different perovskite analogues converge into a single trend, but the data for the $\langle 100 \rangle_c$ slip systems do not. Different creep mechanisms are probably responsible for this behaviour. It is usually assumed that creep with the $\langle 110 \rangle_c \{1\bar{1}0\}_c$ systems is controlled by dislocation glide, whereas creep with the $\langle 100 \rangle_c \{001\}_c$ systems would be controlled by dislocation climb⁵. Our study suggests that magnesium silicate perovskite deforms along $\langle 100 \rangle_o$ (that is $\langle 110 \rangle_c$) slip, so its high-temperature rheological behaviour might be able to be extrapolated from the $CaTiO_3$, $SrTiO_3$, $NaNbO_3$ analogue series, as is the case for the low-temperature hardness¹⁹.

It is only in recent years that seismology has provided us with well-documented information on anisotropy in the lower mantle. Such studies suggest that the bulk of the lower mantle is largely isotropic, except for some portions of the D'' layer. Recently,

Table 2 Dislocation contrast factors

<i>hkl</i>	110	111	112	021	022	023	121	122	123	211	006
C_m	0.087	0.120	0.090	0.053	0.141	0.074	0.110	0.053	0.032	0.056	0.042
C_{calc}	0.095	0.106	0.090	0.054	0.145	0.075	0.113	0.046	0.041	0.063	0.036

Measured, C_m , and calculated, C_{calc} , dislocation contrast factors. The C_m values were obtained by the whole profile fitting procedure¹⁷, and the C_{calc} values correspond to dislocations in Table 3 summed up with the appropriate weight factors. The error in the values of C_m is estimated to be between 10% and 15%.

numerical modelling has been used to examine deformation in the lower mantle associated with downwelling of subducting slabs. It has been suggested that dislocation creep is expected to concentrate in these regions²⁰. This deformation mechanism leads to the development of crystal preferred orientation (CPO), which is a major source of seismic anisotropy. In the absence of experimental data on the plastic deformation of magnesium silicate perovskite, previous studies have focused on the contribution of magnesiowüstite only. The present study shows that dislocation creep can be activated in MgSiO₃ perovskite at high temperature, contrary to what has been reported at room temperature²¹. Our results provide information on the geometry of slip in perovskite that is needed to model CPO in perovskite. However, forward modelling of seismic anisotropy requires the understanding of how strain partitions between the minor, weak phase and the major, hard phase of the magnesiowüstite-perovskite assemblage. This could be achieved through high-pressure/high-temperature deformation experiments such as those described here, or by three-dimensional modelling of plastic deformation of a two-phase polycrystalline aggregate. □

Received 19 December 2003; accepted 5 March 2004; doi:10.1038/nature02472.

1. Ringwood, A. E. Phase transformations and their bearing on the constitution and dynamics of the mantle. *Geochim. Cosmochim. Acta* **55**, 2083–2110 (1991).
2. Poirier, J. P., Peyronneau, J., Gesland, J. Y. & Brebec, G. Viscosity and conductivity of the lower mantle; an experimental study on a MgSiO₃ perovskite analogue, KZnF₃. *Phys. Earth Planet. Inter.* **32**, 273–287 (1983).
3. Beauchesne, S. & Poirier, J. P. Creep of barium titanate perovskite: a contribution to a systematic approach to the viscosity of the mantle. *Phys. Earth Planet. Inter.* **55**, 187–199 (1989).
4. Beauchesne, S. & Poirier, J. P. In search of a systematics for the viscosity of perovskites: creep of potassium tantalate and niobate. *Phys. Earth Planet. Inter.* **61**, 182–198 (1990).
5. Wang, Z. C., Karato, S. & Fujino, K. High temperature creep of single crystal strontium titanate (SrTiO₃)—a contribution to creep systematics in perovskites. *Phys. Earth Planet. Inter.* **79**, 299–312 (1993).
6. Li, P., Karato, S. I. & Wang, Z. High-temperature creep in fine-grained polycrystalline CaTiO₃, an analogue material of (Mg, Fe)SiO₃ perovskite. *Phys. Earth Planet. Inter.* **95**, 19–36 (1996).
7. Wang, Z. C., Dupas-Bruzek, C. & Karato, S. High temperature creep of an orthorhombic perovskite—YAlO₃. *Phys. Earth Planet. Inter.* **110**, 51–69 (1999).
8. Chen, J. H., Weidner, D. J. & Vaughan, M. T. The strength of Mg_{0.9}Fe_{0.1}SiO₃ perovskite at high pressure and temperature. *Nature* **419**, 824–826 (2002).
9. Horiuchi, H., Ito, E. & Weidner, D. J. Perovskite-type MgSiO₃: single-crystal X-ray diffraction study. *Am. Mineral.* **72**, 357–360 (1987).
10. Cordier, P. & Rubie, D. C. Plastic deformation of minerals under extreme pressure using a multi-anvil apparatus. *Mater. Sci. Eng. A* **309**, 38–43 (2001).
11. Caglioti, G., Paoletti, A. & Ricci, F. P. Choice of collimators for a crystal spectrometer for neutron diffraction. *Nucl. Instrum. Methods* **3**, 223–228 (1958).
12. Krivoglaz, M. A. *Theory of X-ray and Thermal Neutron Scattering by Real Crystals* (Plenum, New York, 1969).
13. Krivoglaz, M. A. *X-ray and Neutron Diffraction in Nonideal Crystals* (Springer, Berlin, 1996).
14. Wilkens, M. The determination of density and distribution of dislocations in deformed single crystals from broadened x-ray diffraction profiles. *Phys. Status Solidi* **2**, 359–370 (1970).
15. Kuzel, R. & Klimanek, P. X-ray diffraction line broadening due to dislocations in non-cubic crystalline materials. 3. Experimental results for plastically deformed zirconium. *J. Appl. Crystallogr.* **22**, 299–307 (1989).
16. Ungár, T. & Borbély, A. The effect of dislocation contrast on x-ray line broadening: a new approach to line profile analysis. *Appl. Phys. Lett.* **69**, 3173–3175 (1996).
17. Ungár, T., Gubicza, J., Ribárik, G. & Borbély, A. Crystallite size distribution and dislocation structure determined by diffraction profile analysis: principles and practical application to cubic and hexagonal crystals. *J. Appl. Crystallogr.* **34**, 298–310 (2001).
18. Ribárik, G., Ungár, T. & Gubicza, J. MWP-fit: a program for multiple whole-profile fitting of diffraction peak profiles by ab initio theoretical functions. *J. Appl. Crystallogr.* **34**, 669–676 (2001).
19. Karato, S. I. Plasticity of MgSiO₃ perovskite: the results of microhardness tests on single crystals. *Geophys. Res. Lett.* **17**, 13–16 (1990).
20. McNamara, A. K., van Keken, P. E. & Karato, S. I. Development of anisotropic structure in the Earth's lower mantle by solid-state convection. *Nature* **416**, 310–314 (2002).
21. Merkel, S. et al. Deformation of (Mg_{0.9}Fe_{0.1})SiO₃ perovskite aggregates up to 32 GPa. *Earth Planet. Sci. Lett.* **209**, 351–360 (2003).

Acknowledgements High-pressure experiments were performed at the Bayerisches Geoinstitut under the EU 'IHP—Access to Research Infrastructures' programme. P.C. benefited from a 'Congé thématique pour recherche' from the University of Lille and from continuous support from INSU ('Intérieur de la Terre' programme). T.U. and G.T. thank the Hungarian Science Foundation for supporting this work.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.C. (Patrick.Cordier@univ-lille1.fr).

Competition and mutualism among the gut helminths of a mammalian host

Joanne Lello¹, Brian Boag², Andrew Fenton³, Ian R. Stevenson⁴ & Peter J. Hudson⁵

¹CSIRO Livestock Industries, F.D. McMaster Laboratories, Locked Bag 1, PO Armidale, New South Wales 2350, Australia

²Birch Brae, Knapp, Perth & Kinross PH14 9SW, UK

³Institute of Zoology, ZSL, Regents Park, London NW1 4RY, UK

⁴Department of Biological Sciences, University of Stirling, Stirling FK9 4LA, UK

⁵Department of Biology, 513 Mueller Laboratory, University Park, Penn State University, Pennsylvania 16802, USA

Most animal species are infected with multiple parasite species; however, the role of interspecific parasite interactions in influencing parasite dynamics and shaping parasite communities has been unclear. Although laboratory studies have found evidence of cross-immunity, immunosuppression and competition^{1–6}, analyses of hosts in the field have generally concluded that parasite communities are little more than random assemblages^{7–14}. Here we present evidence of consistent interspecific interactions in a natural mammalian system, revealed through the analysis of parasite intensity data collected from a free-ranging rabbit (*Oryctolagus cuniculus*) population, sampled monthly for a period of 23 yr. The wild rabbit plays host to a diverse gut helminth community^{15–17} that reflects the communities seen in other economically important domestic herbivores^{18,19}. These findings suggest that parasite interactions could have profound implications for the dynamics of parasite communities. The efficacy of parasite control programmes could be jeopardized if such interactions are not taken into account. In contrast, a clear understanding of such interactions may provide the basis for the development of more environmentally acceptable methods of parasite control.

The wild rabbit population in the United Kingdom is dominated by five gut helminths: the strongylid nematodes *Graphidium strigosum* (in the stomach) and *Trichostrongylus retortaeformis* (small intestine); the anoplocephaloid cestodes *Mosgovoyia pectinata* (small intestine) and *Cittotaenia denticulata* (small intestine); and the oxyurid nematode *Passalurus ambiguus* (large intestine and colon). Our objectives were to identify (1) whether interspecific interactions could be detected within this parasite community; (2) quantify the strengths of these interactions; (3) identify the putative mechanisms by which interactions may be mediated; and (4) propose the possible consequences of such interactions for future parasite control.

We used a combination of generalized linear modelling (GLM) and residual maximum likelihood (REML) linear mixed model analyses of parasite count data to test the null hypothesis that parasite interactions do not influence parasite numbers. We included aspects of the host biology and the external environment that if excluded would be most likely to create the appearance of interspecific interactions where none exist (for further details see Supplementary Information). All predictions are presented as the percentage difference between the predicted level of the dependent variable (after back-transformation) when the interacting species has a zero count, compared with the predicted level of the dependent variable (after back-transformation) when the interacting species is at its geometric mean (in concomitant infection).

According to the models, same-locale, downstream (anterior to posterior) and upstream (posterior to anterior) interactions occurred between the helminths. One significant ($P < 0.001$) inter-

action was seen between two species in the same location in the gut, with the model predicting a positive effect of *C. denticulata* on the intensity of *T. retortaeformis* infection (Fig. 1a and Table 1). The geometric mean of *C. denticulata* when in a concomitant infection with *T. retortaeformis* was around two individuals, which was predicted to increase the number of *T. retortaeformis* by 51% relative to a rabbit harbouring no *C. denticulata*.

Trichostrongylus retortaeformis was subject to a predicted, positive downstream influence from *G. strigosum* ($P < 0.001$; Fig. 1b and Table 1). In concomitant infections, the geometric mean of 18 *G. strigosum* was predicted to increase the number of *T. retortaeformis* by 36%. A more complex downstream interaction was seen with *M. pectinata* in the small intestine, which was predicted to have a positive effect in female ($P = 0.003$), but not male, hosts on the intensity of *P. ambiguus* in the large intestine/colon (Fig. 1c and Table 1). The geometric mean of three *M. pectinata* was predicted to increase the intensity of *P. ambiguus* by 163% relative to a host containing no *M. pectinata*.

The model predicted that the association between these two species also operated in the opposite, upstream direction, with *P. ambiguus* having a positive effect on *M. pectinata*. Again, this association was complicated by host sex, such that *P. ambiguus* increased the intensity of *M. pectinata* in male but not female rabbits ($P = 0.008$; Fig. 1d and Table 1). The geometric mean of *P. ambiguus* (in concomitant infection) in male rabbits was 127, which was

predicted to increase *M. pectinata* numbers by 67% relative to hosts not infected with *P. ambiguus*. *Mosgovioya pectinata* also had an upstream effect, reducing the intensity of *G. strigosum* in the stomach ($P = 0.011$; Fig. 1e and Table 1) by 19% on average. Finally, the presence of *T. retortaeformis* was predicted to affect negatively the numbers of *G. strigosum* in an upstream position ($P = 0.018$; Fig. 1f and Table 1). This effect was complicated by host mass (a proxy for age²⁰), such that the effect decreased as rabbit mass increased; however, in a rabbit of average mass (1,590 g in concomitantly infected rabbits), *T. retortaeformis* was predicted to reduce *G. strigosum* numbers by 29%.

Analyses indicate the existence of a network of interactions between these helminths (Fig. 2). In most cases, based on the biology of the parasites, we can postulate the likely mechanisms underlying these interactions. The observed same-locale or downstream interactions may be due to direct influences of one parasite on another through by-products, manipulation of gut physiology, competition or physical crowding. However, for upstream interactions, the most likely route would be interaction mediated through the host's immune system, although changes in gut physiology, for example pH change, cannot be entirely ruled out²¹.

The putative explanation for the observed interactions between *G. strigosum* and *T. retortaeformis* is cross-immunity between these species. Age–intensity curves for *T. retortaeformis* and *G. strigosum* (with mass used as a proxy for age) support previous findings^{22–24}

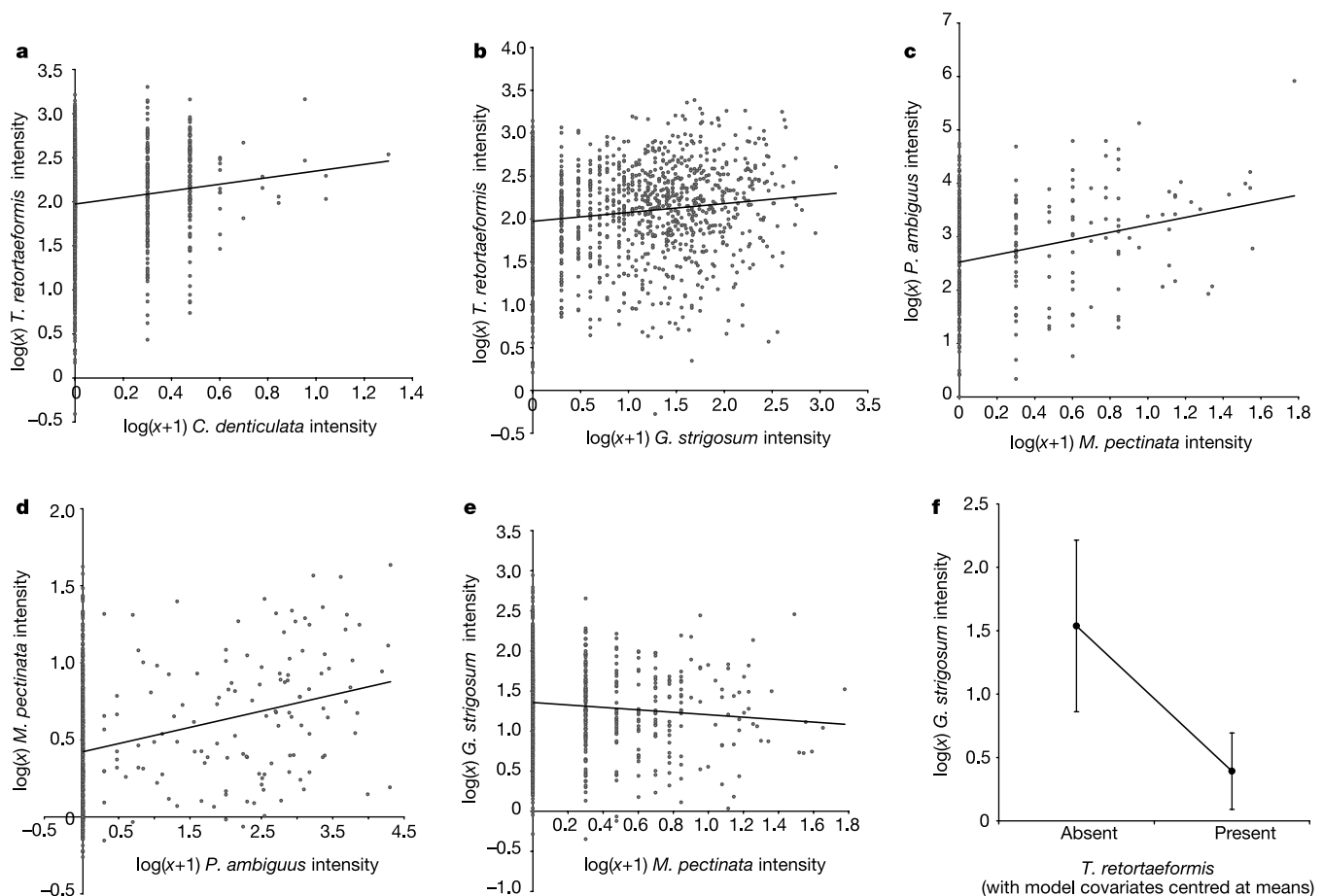


Figure 1 Predictions from generalized linear and restricted maximum likelihood models of the relationships between the dependent variable (affected species $\log(x)$ helminth intensity with zeros removed) and the independent variable **a–e**, effector species $\log(x + 1)$ helminth intensity, or **f**, effector species presence-absence data (in rabbits of average mass, 1,590 g). All analyses were conducted on adult rabbit data. Lines of

predicted fit and residual fit points (or 95% confidence intervals) are shown in each case. *Trichostrongylus retortaeformis* with *C. denticulata* (**a**); *T. retortaeformis* with *G. strigosum* (**b**); *P. ambiguus* with *M. pectinata* (**c**); *M. pectinata* with *P. ambiguus* (**d**); *G. strigosum* with *M. pectinata* (**e**); and *G. strigosum* with *T. retortaeformis* (**f**).

Table 1 Significance of factors affecting the log intensity of gut helminths of *O. cuniculus*

Model term	Test statistic†	d.f.	P value
log <i>G. strigosum</i> intensity‡			
Month	182.50	11	<0.001
Host food type	25.18	1	<0.001
Presence/absence of myxomatosis	8.84	1	0.018
<i>T. retortaeformis</i> *host mass log($x + 1$) <i>M. pectinata</i>	6.6	1	0.011
log <i>M. pectinata</i> intensity§			
Month	62.49	11	<0.001
Sex	0.00	1	0.951
Sex*log($x + 1$) <i>P. ambiguus</i>	7.14	1	0.008
log <i>T. retortaeformis</i> intensity			
Month*host sex	31.5	11,1195	<0.001
Host sex*host mass	2.8	1,1195	0.003
Presence/absence of myxomatosis	12.8	1,1195	<0.001
Log($x + 1$) <i>C. denticulata</i>	23.98	1,1195	<0.001
Log($x + 1$) <i>G. strigosum</i>	31.94	1,1195	<0.001
log <i>C. denticulata</i> intensity			
Month	1.41	11,288	<0.001
log <i>P. ambiguus</i> intensity			
Month	26.4	11,280	0.011
Host sex	1.6	1,280	0.223
Host sex*log($x + 1$) <i>M. pectinata</i>	9.3	1,280	0.003
Warren vegetation class	16.2	1,280	<0.001
Presence/absence of myxomatosis	8.1	1,280	0.011

All effects displayed were significant in the minimal model after bootstrapping (with the *P* value set to 0.05 and iterations = 1,000). d.f., degrees of freedom.

†The test statistic was the Wald Statistic (χ^2) for *G. strigosum* and *M. pectinata*, and the *F* Statistic for *T. retortaeformis*, *C. denticulata* and *P. ambiguus*.

‡Model type: REML (random terms = warren number + year).

§Model type: REML (random term = warren number).

||Model type: GLM.

that *T. retortaeformis* stimulates an acquired immune response whereas the *G. strigosum* numbers show no decline, suggesting no response to host immunity (Supplementary Note and Fig. S1). It is likely that an ability of *G. strigosum*, as seen in other helminths²⁵, to modulate host immunity explains the positive effect this species appears to have on *T. retortaeformis*. Conversely, stimulation of acquired immunity by *T. retortaeformis* may negatively affect the related *G. strigosum*.

The interaction of *P. ambiguus* and *M. pectinata* is affected in both directions by host sex, once again implying mediation through host biology. Similarly, the negative effect of *M. pectinata* on upstream abundance of *G. strigosum* must be host mediated but the mechanism for these interactions is not discernible from the current data. The mechanism for the relationship between the cestode *C. denticulata* and the nematode *T. retortaeformis* cannot be determined, as the current data do not indicate whether the interaction is direct or host mediated.

These analyses provide credible evidence for the presence of interspecific parasite interactions throughout a community of gut parasites, suggesting that host immunity may have a role in shaping that community. Although we wish to stress that the predictions are based on analysis of correlated data and must be validated experimentally, the biology of the helminths provides support that the interactions are genuine.

Notably, there is no reason to believe that the rabbit system is unusual and therefore such analyses applied to other systems, particularly those of economically important livestock (where laboratory experiments have provided evidence of interaction^{21,26}), are likely to reveal similar networks of interactions. The presence and detection of interspecific parasite interactions, particularly where they are immune mediated, could have important implications for parasite control strategies. This is particularly pertinent to the future of helminth control, where rapidly developing resistance to chemotherapeutics is leading to increased efforts in the field of vaccine development²⁷. The work described here suggests that within-host interspecific parasite interactions may have a role in determining the overall impact of such vaccines on gut helminth communities. To illustrate this we used a simple generic model,

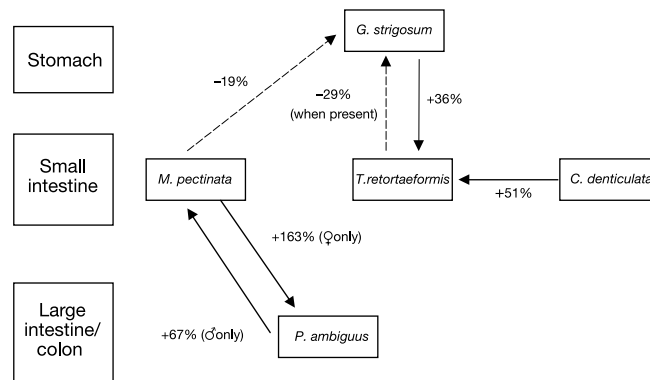


Figure 2 Relative positions of the rabbit gut helminths in the host gut, with interaction directions and strengths, based on observed mean burdens.

which may be adapted to represent any system. The model consisted of three parasite species within a single host (for details see Supplementary Methods). Each parasite stimulated a specific immune response at a rate proportional to the abundance of that parasite, and this immune response acted on the establishment rate of the parasite. Interspecific interactions were incorporated by allowing the immune response against one parasite to affect one of the other species, either increasing or suppressing the action of the immune response against that species. A simple network of interactions between the parasites was constructed such that parasite 1 (P_1) had a negative effect on parasite 2 (P_2), P_2 a positive effect on parasite 3 (P_3), and the effect of P_3 on P_1 (described by the parameter γ_3) was varied from highly negative to highly positive. A species-specific vaccine was then applied, targeted solely against P_1 , and the impact of the strength of interaction between P_3 and P_1 on the efficacy of this vaccine was assessed.

Vaccination against the target parasite always suppressed its abundance below pre-treatment levels ($\Delta_1 < 1$; Fig. 3b). However, vaccination against P_1 tended to have a positive effect on P_2 ($\Delta_2 > 1$) (Fig. 3c), but little impact on P_3 ($\Delta_3 \sim 1$; Fig. 3d), presumably because P_3 was two steps away from the target of the vaccine and was to some extent buffered from its effects.

When γ_3 was 0 (that is, P_3 did not interact with P_1), vaccination reduced P_1 to approximately 15% of its pre-vaccination level (Fig. 3b). Furthermore, as P_1 had a negative interaction with P_2 , its suppression meant that vaccination increased the abundance of P_2 (Fig. 3c). When P_3 had a strong negative interaction with P_1 ($\gamma_3 = -1$) the vaccine, combined with this negative interaction, reduced P_1 abundance to approximately 1% of its pre-vaccination level. As P_1 was at very low abundance, its impact on P_2 was negligible and so P_2 approximated its pre-vaccination level (Δ_2 approached 1). However, when γ_3 was strongly positive (+1) P_1 abundance was only reduced to about 50% of its pre-vaccination level; the positive impact of P_3 on P_1 diminished the efficacy of the vaccine. Notably, given a positive interaction between P_3 and P_1 , vaccinating to control P_1 resulted in a massive relative increase in the abundance of P_2 , up to an order of magnitude greater than its pre-vaccination level. This is surprising because P_1 abundance is increased by the presence of P_3 and so P_2 abundance should be further suppressed due to the negative interaction between P_1 and P_2 . The explanation for this counterintuitive result lies in the subtle nonlinear effects of the interactions. Before vaccination, the positive impact of P_3 on P_1 meant that P_1 was at a very high equilibrium level. Hence, P_2 was suppressed to a very low level. By applying the vaccine, P_1 abundance was reduced to approximately half of its pre-vaccination level, thereby decreasing the negative impact on P_2 . P_2 abundance therefore increased substantially, resulting in a massive relative increase compared with its pre-vaccination level.

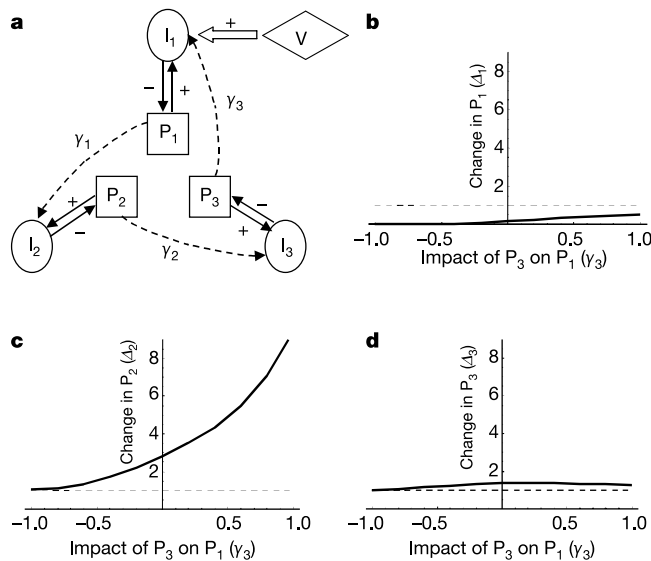


Figure 3 Three species parasite interaction model. **a**, Schematic of three parasite species interaction model (for details see Supplementary Methods) where P_1 , P_2 and P_3 are the parasite species, I_1 , I_2 and I_3 are the species-specific immune responses, V is the vaccine applied against species 1 and γ_1 , γ_2 and γ_3 are the parameters describing the strength of the effect of one species on another. **b–d**, Proportional change (ΔI_i) in equilibrium abundances of parasite species i where $i = 1$ (**b**), $i = 2$ (**c**) and $i = 3$ (**d**), following vaccination (that is, post-vaccination equilibrium level/pre-vaccination equilibrium) at varying levels of γ_3 , the parameter describing the strength of the effect of P_3 on P_1 . The dotted horizontal line represents where the post-vaccination level = pre-vaccination level.

Hence, vaccination against one parasite species produced unexpected and marked increases in the abundance of another, previously innocuous parasite species, by releasing it from the negative impact of the targeted parasite. These, and other, important effects would be missed if we ignored the possibility of such interactions.

The models used here have revealed an array of potential interactions between parasites, which would not have been elucidated by less powerful techniques. As concomitant infections are the norm rather than the exception²⁸, application of these techniques to other mammalian-parasite data sets is likely to reveal similar interactions. Given the increase of anthelmintic resistance among gut helminths²⁹, alternative control strategies are urgently needed. A clearer understanding of parasite community ecology may provide such an alternative. One implication is that parasites could be used to control other infections. For example, artificial infection with one relatively benign species could be used to stimulate cross-immunity and prevent subsequent infection with a deleterious species without resorting to anthelmintics. However, without the knowledge of the community relationships, treatment against one deleterious species could allow an increase in the intensity of a second harmful species. Previously unexplored interactions may be one explanation for the failure of laboratory proven vaccines to work under more natural conditions. It is clear that such interactions need to be incorporated into future dynamical studies of parasite communities where (with a few exceptions³⁰) parasite interspecific conflicts and mutualisms are ignored in the null models. □

Methods

Rabbits were collected from a 400 ha site in Perthshire, Scotland (ordnance grid reference NO 280 340) between the months of January 1977 and December 1999. Total gut helminth counts, presence of myxomatosis and presence of *Eimeria stiedae* were recorded along with details of the external environment and aspects of host biology. Further details of the study site, collection protocols and data may be found in ref. 17.

All analyses were conducted using the GENSTAT statistical package (VSN International Ltd) and all models were initially run as REMLs with the random terms of warren code and

year of capture. Interactions between all model terms were initially included up to third order. Only adult rabbits (mass >1,249 g; $n = 1,526$) were used for this analysis because they were likely to have been exposed to infective stages of all parasites by this age. In addition, the distributions of the gut helminth parasites in different age classes of the rabbit have been shown to be different and must therefore be analysed separately¹⁷. The data for the gut helminths do not conform to the negative binomial distribution and in order to obtain a 'near' normal distribution for the dependent variable, zeros had to be removed from the data before a log transformation was applied¹⁷. We therefore considered those factors that affected the intensity of the species treated as the dependent variable (in infected animals only). Each helminth species in turn was treated as the dependent variable. The helminths not being treated as the dependent variable were included in the model (as determined from submodels) as independent factors (presence/absence) or as independent variables ($\log(x + 1)$ intensity data) as appropriate. Nonsignificant terms were removed from the models in a stepwise manner until a minimal model was produced. This model was further refined following bootstrapping of the model effects and removal of any terms that were not significantly different from zero. Once the bootstrapping refinements were completed, each model was run once again in its final form to obtain effect sizes and significance levels. As model terms were only retained after the bootstrap of their effect size, $\log(x + 1)$ transformation was considered sufficient for the species treated as independent variables and allowed comparison of concomitantly infected rabbits with those only infected with the species treated as the dependent variable.

Received 18 December 2003; accepted 10 March 2004; doi:10.1038/nature02490.

- Behnke, J. M., Wakelin, D. & Wilson, M. M. *Trichinella spiralis*: Delayed rejection in mice concurrently infected with *Nematostomus dubius*. *Exp. Parasitol.* **46**, 121–130 (1978).
- Silver, B. B., Dick, T. A. & Welch, H. E. Concurrent infections of *Hymenolepis diminuta* and *Trichinella spiralis* in the rat intestine. *J. Parasitol.* **66**, 786–791 (1980).
- Christensen, N. O., Nansen, P., Fagbemi, B. O. & Monrad, J. Heterologous antagonistic and synergistic interactions between helminths and between helminths and protozoans in concurrent experimental-infection of mammalian hosts. *Parasitol. Res.* **73**, 387–410 (1987).
- Adams, D. B., Anderson, B. H. & Windon, R. G. Cross-immunity between *Haemonchus contortus* and *Trichostrongylus colubriformis* in sheep. *Int. J. Parasitol.* **19**, 717–722 (1989).
- Behnke, J. M. *et al.* Variation in the helminth community structure in bank voles (*Clethrionomys glareolus*) from three comparable localities in the Mazury Lake District region of Poland. *Parasitology* **123**, 401–414 (2001).
- Cox, F. E. G. Concomitant infections, parasites and immune responses. *Parasitology* **122**, S23–S38 (2001).
- Bush, A. O. & Holmes, J. C. Intestinal helminths of lesser scaup ducks: an interactive community. *Can. J. Zool.* **64**, 142–152 (1986).
- Haukisalmi, V. & Henttonen, H. Coexistence in helminths of the bank vole *Clethrionomys glareolus*. 1. Patterns of cooccurrence. *J. Anim. Ecol.* **62**, 221–229 (1993).
- Haukisalmi, V. & Henttonen, H. Coexistence in helminths of the bank vole *Clethrionomys glareolus*. 2. Intestinal distribution and interspecific interactions. *J. Anim. Ecol.* **62**, 230–238 (1993).
- Forbes, M., Weatherhead, P. J. & Bennett, G. F. Blood parasites of blue grouse—Variation in prevalence and patterns of interspecific association. *Oecologia* **97**, 520–525 (1994).
- Poulin, R. Richness, nestedness, and randomness in parasite infracommunity structure. *Oecologia* **105**, 545–551 (1996).
- Poulin, R. Interactions between species and the structure of helminth communities. *Parasitology* **122**, S3–S11 (2001).
- Nilssen, A. C., Haugerud, R. E. & Folstad, I. No interspecific covariation in intensities of macroparasites of reindeer, *Rangifer tarandus* (L.). *Parasitology* **117**, 273–281 (1998).
- Forbes, M. R., Alisauskas, R. T., McLaughlin, J. D. & Cuddington, K. M. Explaining co-occurrence among helminth species of lesser snow geese (*Chen caerulescens*) during their winter and spring migration. *Oecologia* **120**, 613–620 (1999).
- Mead-Briggs, A. R. & Vaughan, J. A. The incidence of anoplocephaline cestodes in a population of rabbits in Surrey, England. *Parasitology* **67**, 351–364 (1973).
- Mead-Briggs, A. R. & Page, R. J. C. Records of anoplocephaline cestodes from wild rabbits and hares collected throughout Great Britain. *J. Helminthol.* **49**, 49–56 (1975).
- Boag, B., Lello, J., Fenton, A., Tompkins, D. M. & Hudson, P. J. Patterns of parasite aggregation in the wild European rabbit (*Oryctolagus cuniculus*). *Int. J. Parasitol.* **31**, 1421–1428 (2001).
- Crofton, H. D. The ecology of the immature phases of trichostrongyle nematodes. v. The estimation of pasture infections. *Parasitology* **44**, 313–324 (1954).
- Boag, B. Similarities between the population dynamics of parasites of the wild rabbit and domestic animals. *Asp. Appl. Biol.* **53**, 119–123 (1999).
- Cowan, D. P. *Aspects of the Behaviour of a Free-Living Population of the European Wild Rabbit, Oryctolagus cuniculus L. in Southern England*. Thesis, Royal Holloway College, Univ. London (1983).
- Dobson, R. J. & Barnes, E. H. Interaction between *Ostertagia circumcincta* and *Haemonchus contortus* infection in young lambs. *Int. J. Parasitol.* **25**, 495–501 (1995).
- Michel, J. F. Phenomenon of protection in infections of *Trichostrongylus retortaeformis*. *Nature* **172**, 312 (1955).
- Ford, G. E. Heterogeneity of allergens from, and homocytotropic antibody to, a gastro-intestinal nematode of rabbits. *Immunology* **21**, 1073–1078 (1971).
- Boag, B. & Kolb, H. H. Influence of host age and sex on nematode populations in the wild rabbit (*Oryctolagus cuniculus* L.). *Proc. Helminthol. Soc. Wash.* **56**, 116–119 (1989).
- Behnke, J. M., Bajer, A., Sinski, E. & Wakelin, D. Interactions involving intestinal nematodes of rodents experimental and field studies. *Parasitology* **122**, S39–S49 (2001).
- Christensen, C. M., Nansen, P. & Barnes, E. H. The effect of concurrent or sequential *Oesophagostomum dentatum* and *O. quadrispinulatus* infections on the worm burdens of the two species in pigs. *Parasitology* **114**, 273–278 (1997).
- Dalton, J. P. *et al.* Helminth vaccines: from mining genomic information for vaccine targets to systems used for protein expression. *Int. J. Parasitol.* **33**, 621–640 (2003).
- Petney, T. N. & Andrews, R. H. Multiparasite communities in animals and humans: frequency, structure and pathogenic significance. *Int. J. Parasitol.* **28**, 377–393 (1998).

29. Cornell, S. J., Isham, V. S., Smith, G. & Grenfell, B. T. Spatial parasite transmission, drug resistance, and the spread of rare genes. *Proc. Natl Acad. Sci. USA* **100**, 7401–7405 (2003).
30. Dobson, A. P. & Roberts, M. The population dynamics of parasitic helminth communities. *Parasitology* **109**, S97–S108 (1994).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This work was undertaken at the University of Stirling and we thank the Natural Environment Research Council for funding the project. We also thank R. Norman of the University of Stirling for modelling assistance.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.L. (Joanne.Lello@csiro.au).

Tug-of-war over reproduction in a social bee

Philipp Langer¹, Katja Hogendoorn^{2*} & Laurent Keller¹

¹Department of Ecology and Evolution, University of Lausanne, Biology Building, 1015 Lausanne, Switzerland

²School of Biological Sciences, Flinders University of South Australia, Adelaide SA 5001, Australia

* Present address: Centre for Plant and Pest Management, Adelaide University, Glen Osmond SA 5064, Australia

One of the main transitions in evolution is the shift from solitary organisms to societies with reproductive division of labour^{1,2}. Understanding social evolution requires us to determine how ecological, social and genetic factors jointly influence group stability and partitioning of reproduction between group members^{3–8}. Here we test the role of the three key factors predicted to influence social evolution by experimentally manipulating them in a social allopapine bee. We show that increased relatedness between nestmates results in more even reproduction among group members and a greater productivity per individual. By contrast, the degree of reproductive skew is not influenced by the opportunity for solitary breeding or by the potential benefits of cooperation. Relatedness also has a positive effect on group stability and overall productivity. These findings are in line with predictions of the tug-of-war models, in which the degree of reproductive division of labour is determined primarily by selfish competition between group members. The alternative view, where the degree of reproductive skew is the outcome of a social contract between potential breeders, was not supported by the data.

Reproductive skew models can be divided into two classes on the basis of the assumed power relations in a social group. In transactional models, group members yield reproduction to each other in return for specific benefits, whereas in tug-of-war models reproductive sharing reflects each group member's inability to effectively monopolize reproduction^{5,6} (Table 1). The first of the two main transactional models, the concession model, assumes that a dominant individual fully controls both group membership and the fraction of total group reproduction that a subordinate obtains^{9–11}. Under this model, the dominant yields just enough reproduction to a subordinate to make it worthwhile for the subordinate to stay and cooperate peacefully rather than leave the group to reproduce solitary. This model predicts that reproductive skew increases with higher genetic relatedness between the dominant and the subordinate (r), and higher overall reproductive output of a group of two breeders relative to the output of a single dominant

(k), but decreases with higher expected solitary reproduction of a potential subordinate relative to that of a lone dominant (x , lower values of x indicating harsher ecological constraints on solitary breeding). This is because the subordinate gains greater fitness benefits from cooperation with increased r and k , and thus requires less direct reproduction. However, when the chances of successful solitary breeding of the subordinate are high (increased x), the dominant has to concede a larger fraction of colony reproduction to the subordinate to make staying in the group profitable for the latter. The other main transactional model, the restraint model, assumes that the dominant controls group membership but that a subordinate fully controls its reproductive share within the group¹². Under this model, the subordinate captures the largest share of reproduction that the dominant will tolerate before ejecting the subordinate, resulting in opposite predictions to the concession model (Table 1). Finally, tug-of-war models assume that both dominant and subordinate individuals have only limited control over the allocation of reproduction and must expend effort to increase their shares of the total group output. The main prediction of this model is that reproductive skew is negatively correlated with or independent of relatedness because struggle over reproduction takes place at the cost of overall group productivity, with the effect that higher relatedness acts as a break on the investment in reproductive competition¹³. This model predicts that skew is not correlated with group productivity (k) or the ecological constraint on solitary breeding (x). Unlike the two transactional models, where relatedness has no influence on group productivity, the tug-of-war model predicts a positive correlation between breeder relatedness and total group output, because related females allocate less energy to reproductive competition.

We tested the predictions of reproductive skew models in the Australian allopapine bee *Exoneura nigrescens* by experimentally manipulating all three parameters (r , x and k). Several characteristics make this species an ideal experimental system. First, cooperation is facultative, with females nesting either solitary or with up to three other females¹⁴, which allows direct comparison between social and solitary productivities (k). Second, there are no morphological castes, hence all females have the potential to reproduce¹¹. Third, a high proportion (51%) of multi-female nests contain two females¹⁵, the type of groups for which most reproductive skew models have been developed. Fourth, within-nest relatedness (r) is variable^{14,16,17} and can be experimentally manipulated. Fifth, nests are built primarily in dead dried flower stalks of grasstrees (*Xanthorrhoea minor*) and *Melaleuca squarrosa*¹⁵,

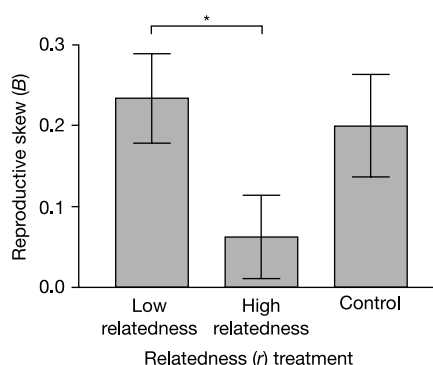


Figure 1 Reproductive skew in nests of the high- and low-relatedness treatment, and in unmanipulated control nests. Overall skew was significantly greater than zero in both the low-relatedness treatment (one-sample t -test, $t = 4.2$, d.f. = 6, $P = 0.005$) and natural nests ($t = 3.2$, d.f. = 8, $P = 0.01$). In the high-relatedness treatment, however, skew was not significantly different from zero ($t = 1.2$, d.f. = 12, $P = 0.24$), the expected value if breeders share reproduction equally²⁹. Error bars indicate standard errors, $n = 29$.

Table 1 Observed and predicted association between variables according to different families of skew models^a

Association	Concession models	Restraint model	Tug-of-war model	Experimental data
Skew versus relatedness (<i>r</i>)	Positive	Negative	None or negative	Negative
Skew versus subordinate's solitary nesting success (<i>x</i>)	Negative	Positive	None	None
Skew versus productivity benefits of cooperative nesting (<i>k</i>)	Positive	Negative	None	None
Relatedness (<i>r</i>) versus overall group productivity	None	None	Positive	Positive

The effect of each parameter assumes that the other parameters and group size are held constant.

providing opportunities for controlling nesting site abundance and thus *x*, the constraints on solitary breeding. Sixth, because nesting site abundance varies greatly in the wild^{17,18}, individuals are likely to be responsive to experimental manipulation of nesting opportunities. Finally, whole colonies can be easily transferred to experimental plots that differ in specific resources, allowing the manipulation of *k*, the relative productivity of social versus solitary nests.

To test the role of relatedness (*r*) on reproductive skew, nest confounding by related bees was favoured in a high-relatedness treatment but restrained in a low-relatedness treatment (see Methods). The treatment proved successful, resulting in a significantly greater relatedness among foundresses in the high-relatedness treatment (mean $r \pm \text{s.e.m.} = 0.51 \pm 0.08$) compared with the low-relatedness treatment ($r = 0.08 \pm 0.08$; unpaired *t*-test, $t = 3.60$, d.f. = 19, $P < 0.002$), whereas unmanipulated control nests showed an intermediate value ($r = 0.38 \pm 0.11$). Relatedness had a significant effect on the degree of reproductive skew, with significantly higher skew (unpaired *t*-test, $t = 2.11$, d.f. = 18, $P < 0.05$) in the low-relatedness compared with the high-relatedness treatment (Fig. 1). The influence of relatedness on skew was further demonstrated by a significant negative correlation between skew and relatedness across all colonies (Pearson's $r = -0.53$, $n = 29$, $P < 0.005$) (Fig. 2). Variation in relatedness also affected overall group productivity: two-female nests in the high-relatedness treatment (7.2 ± 0.9 offspring) were significantly more productive (unpaired *t*-test, $t = 2.31$, d.f. = 17, $P = 0.03$) than two-female nests in the low-relatedness treatment (4.0 ± 0.4 offspring). Across all colonies, the productivity per female was also positively correlated with breeder relatedness (Pearson's $r = 0.35$, $n = 32$, $P < 0.05$).

To test the influence of solitary breeding success (*x*) on reproductive skew, we manipulated the availability of nesting sites. In a low-nest-availability treatment we removed all nesting substrates (dead stalks of grasses), whereas in a high-nest-availability treatment we added supplementary nesting sites. Our treatment was clearly effective, because females in the low-nest-availability treatment, lacking alternative nesting substrate, remained in their initial nest in a significantly higher proportion (26 out of 66 nests) than did those in the high-nest-availability treatment (11 out of 78 nests; Fisher's exact test, $P = 0.001$). However, nest availability had

no significant effect on reproductive skew (unpaired *t*-test, $t = 0.39$, d.f. = 14, $P = 0.70$) (Fig. 3).

To test the influence of *k*, the relative productivity of two- versus one-female nests, we placed occupied nests in plots with rich and poor floral resources, because this has been shown to affect the relative benefits of group living in this species¹⁹. Our treatment was successful given that two-female nests were disproportionately more productive than single-female nests in rich floral plots ($k = 3.4$), but not in poor floral plots where single- and two-female nests showed similar productivities ($k = 1.1$). Consequently, a two-way ANOVA on productivity revealed a significant interaction between the presence of one versus two females per nest and the level of floral resources ($F = 10.1$, d.f._{n,d} = 1,88, $P = 0.002$); the number of females per nest ($F = 12.5$, d.f._{n,d} = 1,88, $P < 0.001$) and the level of floral resources ($F = 10.4$, d.f._{n,d} = 1,88, $P < 0.002$) also had a significant effect on productivity. Despite the important difference in *k* between rich and poor floral plots, there was no significant difference in skew (unpaired *t*-test, $t = 0.86$, d.f. = 22, $P = 0.40$) (Fig. 4). It is unlikely that the lack of difference in skew can be explained by the inability of females to distinguish between the two environments, because the bees founded significantly more joint colonies in rich floral plots, the environment providing larger rewards for cooperation (20 multi-female and 22 single-female nests in rich floral plots versus 14 multi-female and 47 single-female nests in poor floral plots; two-sided Fisher's exact test, $P = 0.01$).

The results of this study do not support the predictions of the two transactional models (Table 1). None of the four predictions of the concession model was found to hold in *E. nigrescens*. Similarly, three predictions of the restraint model were not supported by our data. There was no significant positive relationship between reproductive skew and the constraints on solitary breeding (*x*), and no significant negative relationship between skew and the relative productivity of two- versus one-female nests (*k*). Importantly, the lack of significant association between skew and relative productivity is unlikely to result from a lack of power of the test, because the restraint model would predict complete skew in the rich floral plots (calculated from equation (2) in ref. 12 using the observed parameters, $k = 3.4$ and $r = 0.52 \pm 0.08$, and all possible values of *e*, the cost of eviction, and *x*), whereas skew was far from complete in the experimental data. Finally, further evidence against the restraint model comes

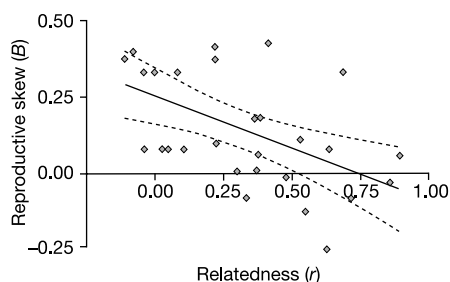


Figure 2 Correlation between relatedness and reproductive skew across all nests. The data include unmanipulated natural nests as well as high-relatedness and low-relatedness treatment nests. The straight line indicates the linear regression with its 95% confidence interval (dotted curve), slope -0.34 ± 0.10 , $n = 29$.

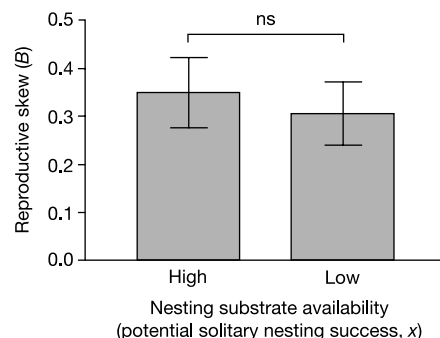


Figure 3 Reproductive skew in treatments with high or low nesting substrate availability. The availability of nesting substrate is a direct measure of a nestmate's potential solitary nesting success (*x*). Error bars indicate standard errors, $n = 16$.

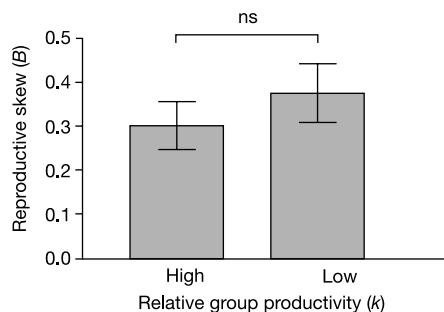


Figure 4 Reproductive skew in treatments with rich and poor floral resources (high and low k , respectively). Error bars indicate standard errors, $n = 24$.

from the finding that group productivity was positively associated with relatedness, whereas this model predicts no association.

By contrast, the results matched each of the four predictions of the tug-of-war models. Reproductive skew was negatively associated with the relatedness between breeders (r) and independent of both the constraints on solitary breeding (x) and the productivity benefits of cooperative nesting (k). Moreover, overall group productivity was positively associated with r , as predicted by tug-of-war models. This association, which has previously been reported in some other social insects²⁰, is expected from the lower investments in reproductive competition in groups where individuals are more related because of kin benefits. Finally, we also found that the tendency of females to join another female was significantly influenced by their relatedness: in the high-relatedness treatment, where females could readily associate with related individuals, the bees founded 33 multi-female and 30 single-female nests, versus 5 multi-female and 52 single-female nests in the low-relatedness treatment (two-sided Fisher's exact test, $P < 0.0001$). This result shows that females can assess relatedness and prefer to breed alone rather than with unrelated individuals, despite much lower survival rates of solitary nests in this species²¹. This is in line with tug-of-war models that predict higher productivity and lower competition in nests of related breeders, but is incompatible with transactional models that predict that females should join independently of relatedness^{11,22}.

This study is the first to succeed in experimentally manipulating relatedness between breeders in a natural environment. Importantly, our data show that *Exoneura* females can assess each of the three parameters involved in reproductive skew models (r , x and k) and actively adapt their reproductive strategy accordingly, a basic yet never tested assumption of skew models²³. The overall results strongly support the tug-of-war model, in which a dominant is unable to fully control the allocation of reproduction and group membership in an animal society^{13,24}, and partitioning of reproduction thus results from selfish and costly efforts of individuals to secure the greatest possible share of the group's output. Accordingly, removals of brood and mutual egg eating have both been observed in nests of this species¹⁹. Inversely, there was no support for the alternative view that partitioning of reproduction ultimately reflects a 'social contract' between group members. More generally, this study emphasizes the relative complexity of insects' behavioural repertoire and their ability to adaptively modulate their level of competitive effort according to social attributes such as the relatedness of other group members. □

Methods

Manipulation of relatedness (r)

Four hundred nests were collected in March 1999 at Cobboboonee State Forest, Victoria, Australia. To construct high-relatedness nests, we placed ten nestmate females (related¹⁴) from overwintered nests into each of 30 vials deposited in two 30 × 30 m experimental plots. Within half a metre around each vial, we placed ten stalks of nesting substrate (dry grassflower flower scapes)¹⁸ to provide females with opportunities to initiate a nest with related individuals^{15,17}. To construct low-relatedness nests, we followed the same

procedure but placed ten females from ten different nests in each of 30 vials, which decreased the opportunity for females to initiate a colony with a related female. As for the high-relatedness treatment, the vials were placed in two experimental plots. All nests were collected in June after the founding period and X-rayed to determine nest contents. The multi-female nests were immediately returned to a single field site to provide a similar environment and thus control other ecological parameters possibly influencing skew. Nests were collected in January 2000 after brood production. The experiment was repeated with another 400 nests the following year, swapping the plots used for high- and low-relatedness treatments. As t -tests showed no significant effects of year and plot on productivity, within-nest relatedness or reproductive skew, we pooled the data and present the combined results.

Maternity and relatedness within nests were estimated using microsatellites and dissection data (see below). Three out of 33 nests contained a mother and her daughter as breeders. They were excluded because of the potential effect of matrilineal structure on skew^{13,25}.

Manipulation of ecological constraints on solitary nest founding (x)

In January 2000, we removed all existing nests and nesting substrate in four similar plots of 50 × 50 m, and deposited approximately 40 established nests from a nearby population per plot. In two plots (high-nest-availability treatment), we placed ten stalks of empty nesting substrate nearby each deposited nest, providing ample possibilities for solitary nesting¹⁸. In the other two plots (low-nest-availability treatment), no nesting substrate was provided, resulting in harsh ecological constraints on solitary nesting. All nests were collected in December 2000 for genetic analysis. Because pairs of plots within each treatment showed no significant differences in productivity, within-nest relatedness and reproductive skew (t -tests, all nonsignificant), data were combined.

Manipulation of relative group productivity (k)

A previous study showed that the productivity of multi-female colonies relative to single-female colonies (k) is lower in poor-resource than in high-resource areas¹⁹. To determine the effect of k on skew, we selected three pairs of plots (50 × 50 m each), one with high and the other with low flower abundance. None of the plots originally contained nesting substrates or occupied nests. In July 2000, we placed about 50 newly founded nests in each plot. We collected all nests in December 2000 and assessed productivity of single- and multi-female nests as well as relatedness, maternity and skew. In each plot, the relative productivity of two- versus single-female nests (k) was calculated by dividing the mean number of offspring in two-female nests by the mean number of offspring in single-female nests⁶. The number of offspring did not differ significantly between plots of the same treatment in single-female nests (t -tests, all nonsignificant). Similarly, there was no significant difference in productivity between plots in two-female nests (t -tests, all nonsignificant), and we therefore combined the plots of the same treatment. Two nests containing a breeding mother–daughter association were excluded from the analyses.

Genetic analyses

All adults, brood and eggs were genotyped at five highly polymorphic microsatellite loci (N22, N60, N81, N83, R74; ref. 26). Adults were furthermore dissected, and the sizes of oocytes, insemination status, yellow body content and wing wear recorded²⁷. The genetic and morphological data allowed us to unambiguously distinguish between the foundresses and their daughters, and to determine maternity of all offspring. Sample sizes were 74 females and 335 offspring, 41 females and 220 offspring, and 58 females and 201 offspring for the first, second and third experiment, respectively.

Relatedness between foundresses was calculated using the program Relatedness 5.0.8 (ref. 28). Allele frequencies were bias-corrected at the nest level because nestmates are related, and nests were weighted equally.

Reproductive skew

The degree of reproductive skew was quantified with the B index, where zero indicates random distribution of reproduction among group members, positive values a higher skew and negative values a lower skew than expected by chance²⁹. We quantified skew for overall offspring production because of the low number of males produced. Separate analyses for each sex gave similar qualitative results, as there was no trade-off among breeders for male and female production³⁰. Of the multi-female nests considered in this study, 63 nests contained two females and 20 nests three or more females. Statistical analyses considering only two-female nests or all nests with multiple females yielded identical conclusions. We therefore present only the analyses comprising all nests.

Received 4 December 2003; accepted 23 February 2004; doi:10.1038/nature02431.

- Maynard Smith, J. & Szathmáry, E. *The Major Transitions in Evolution* (W. H. Freeman, Oxford, 1995).
- Keller, L. (ed.) *Levels of Selection in Evolution* (Princeton Univ. Press, Princeton, 1999).
- Keller, L. & Reeve, H. K. Partitioning of reproduction in animal societies. *Trends Ecol. Evol.* **9**, 98–102 (1994).
- Field, J., Solis, C. R., Queller, D. C. & Strassmann, J. E. Social and genetic structure of paper wasp cofoundress associations: tests of reproductive skew models. *Am. Nat.* **151**, 545–563 (1998).
- Johnstone, R. A. Models of reproductive skew: a review and synthesis. *Ethology* **106**, 5–26 (2000).
- Reeve, H. K. & Keller, L. Tests of reproductive-skew models in social insects. *Annu. Rev. Entomol.* **46**, 347–385 (2001).
- Clutton-Brock, T. H. *et al.* Cooperation, control, and concession in meerkat groups. *Science* **291**, 478–481 (2001).
- Haydock, J. & Koenig, W. D. Reproductive skew in the polygynandrous acorn woodpecker. *Proc. Natl Acad. Sci. USA* **99**, 7178–7183 (2002).
- Vehrencamp, S. L. A model for the evolution of despotic versus egalitarian societies. *Anim. Behav.* **31**, 667–682 (1983).
- Emlen, S. T. The evolution of helping. I. An ecological constraints model. *Am. Nat.* **119**, 29–39 (1982).

11. Reeve, H. K. & Ratnieks, F. L. W. in *Queen Number and Sociality in Insects* (ed. Keller, L.) 45–85 (Oxford Univ. Press, Oxford, 1993).
12. Johnstone, R. A. & Cant, M. A. Reproductive skew and the threat of eviction: a new perspective. *Proc. R. Soc. Lond. B* **266**, 275–279 (1999).
13. Reeve, H. K., Emlen, S. T. & Keller, L. Reproductive sharing in animal societies: reproductive incentives or incomplete control by dominant breeders? *Behav. Ecol.* **9**, 267–278 (1998).
14. Schwarz, M. P., Bull, N. J. & Hogendoorn, K. Evolution of sociality in the allodapine bees: a review of sex allocation, ecology and evolution. *Insectes Soc.* **45**, 349–368 (1998).
15. Silberbauer, L. X. & Schwarz, M. P. Life-cycle and social-behavior in a heathland population of the allodapine bee, *Exoneura bicolor* (Hymenoptera, Apidae). *Insectes Soc.* **42**, 201–218 (1995).
16. Schwarz, M. P. Female-biased sex-ratios in a facultatively social bee and their implications for social evolution. *Evolution* **48**, 1684–1697 (1994).
17. Silberbauer, L. X. Founding patterns of *Exoneura bicolor* in Cobboboonee State Forest, S.W. Victoria. *Aust. Zool.* **28**, 67–73 (1992).
18. Bull, N. J. & Schwarz, M. P. The habitat saturation hypothesis and sociality in an allodapine bee: cooperative nesting is not 'making the best of a bad situation'. *Behav. Ecol. Sociobiol.* **39**, 267–274 (1996).
19. Stevens, M. I. *Egg Stock Piling Strategies and Benefits to Group Living in Exoneura bicolor Smith (Apidae, Xylocopinae)* Thesis, Flinders Univ. South Australia, Adelaide (1997).
20. Gadagkar, R., Chandrashekar, K., Chandran, S. & Bhagavan, S. Queen success is correlated with worker–brood genetic relatedness in a primitively eusocial wasp (*Ropalidia marginata*). *Experientia* **49**, 714–717 (1993).
21. Hogendoorn, K. & Zammitt, J. Benefits of cooperative breeding through increased colony survival in an allodapine bee. *Insectes Soc.* **48**, 392–397 (2001).
22. Reeve, H. K. in *The Social Biology of Wasps* (eds Ross, K. G. & Matthews, R. W.) 99–148 (Cornell Univ. Press, Ithaca, 1991).
23. Kokko, H. Are reproductive skew models evolutionarily stable? *Proc. R. Soc. Lond. B* **270**, 265–270 (2003).
24. Clutton-Brock, T. H. Reproductive skew, concessions and limited control. *Trends Ecol. Evol.* **13**, 288–292 (1998).
25. Reeve, H. K. & Keller, L. Partitioning of reproduction in mother–daughter versus sibling associations—a test of optimal skew theory. *Am. Nat.* **145**, 119–132 (1995).
26. Langer, P., Molbo, D. & Keller, L. Polymorphic microsatellite loci in Allodapine bees for investigating the evolution of social behaviour. *Mol. Ecol. Notes* (in the press).
27. Schwarz, M. P. Persistent multi-female nests in an Australian Allodapine bee, *Exoneura bicolor* (Hymenoptera, Anthophoridae). *Insectes Soc.* **33**, 258–277 (1986).
28. Queller, D. C. & Goodnight, K. F. Estimating relatedness using genetic markers. *Evolution* **43**, 258–275 (1989).
29. Nonacs, P. Measuring and using skew in the study of social behavior and evolution. *Am. Nat.* **156**, 577–589 (2000).
30. Fournier, D. & Keller, L. Partitioning of reproduction among queens in the Argentine ant, *Linepithema humile*. *Anim. Behav.* **62**, 1039–1045 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank C. Roger, P. Jordan and G. Deppieraz for help in genotyping, N. DiMarco for extractions, M. P. Schwarz for methodological advice, A. Marazzi for statistical assistance, J. Zammitt, C. van der Muren, E. Geertsema and T. Taraldsrud for help in setting up field experiments and R. Hammond, M. Chapuisat, P. Christe, S. Helms Cahan, K. Parker, D. Queller and F. Ratnieks for comments on the manuscript.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.L. (Philipp.Langer@ie-zea.unil.ch) or L.K. (Laurent.Keller@ie-zea.unil.ch).

Genetic changes associated with floral adaptation restrict future evolutionary potential

Rebecca A. Zufall^{1,2} & Mark D. Rausher¹

¹Department of Biology, Box 90338, Duke University, Durham, North Carolina 27708, USA

²Current address: Department of Biological Sciences, Smith College, Northampton, Massachusetts 01063, USA

A commonly accepted evolutionary principle is that adaptive change constrains the potential directions of future evolutionary change^{1–3}. One manifestation of this is Dollo's law, which states that character elimination is irreversible^{4,5}. Although the common occurrence of irreversibility has been documented by

phylogenetic analyses of phenotypic transitions, little is known about the underlying causes of this phenomenon⁴. One explanation for evolutionary irreversibility relies on the fact that many characteristics result from interactions between multiple gene products^{4,6}. Such characteristics may often be eliminated by inactivation of just one gene in the network. If they serve no other functions, other genes of the network are then free to accumulate mutations or evolve new functions. Evolutionary change after character loss results in the accumulation of redundant loss-of-function mutations. Such pathway degeneration makes it very unlikely that the characteristic will re-evolve, because multiple simultaneous mutations would be required⁴. Here we describe what appear to be the initial stages of such degeneration in the anthocyanin pigment pathway associated with an adaptive change from blue to red flowers in the morning glory *Ipomoea quamoclit*.

The ancestral floral colour in the genus *Ipomoea* is blue/purple^{7,8} (Fig. 1a, b). A number of independent transitions to other colours have occurred in this genus^{7–9}. One such transition is represented by a small, well-supported clade¹⁰ of red-flowered species in the section *Mina*, subgenus *Quamoclit*, including *I. quamoclit*⁷ (Fig. 1a, b), which are pollinated primarily by hummingbirds^{8,11,12}. Unlike typical blue-flowered *Ipomoea* species, which are pollinated primarily by bees and exhibit a set of characteristics associated with adaptation to bee pollination (for example, a broad floral tube, small quantities of nectar, inserted stigma and anthers and non-versatile anthers)^{8,13}, *Mina* species exhibit a suite of floral traits, including red pigmentation, typically associated with bird pollination (for example, narrow tube, copious nectar production, exerted stigma and anthers and versatile anthers)^{7,8,13}. The transition to red flowers in this group appears to represent an adaptive shift to association with a different type of pollinator.

Floral colour in *Ipomoea* is determined largely by the type of anthocyanin pigment produced. Pigments derived from cyanidin typically produce blue/purple flowers, whereas pigments derived from pelargonidin typically produce red flowers^{14–16}. Cyanidin differs from pelargonidin by possessing an extra hydroxyl group, which is added by the enzyme flavonoid 3'-hydroxylase (F3'H)¹⁶. This enzyme creates a branch in the biosynthetic pathway, with the enzymes downstream of F3'H (that is, dihydroflavonol reductase (DFR), anthocyanidin synthase (ANS) and UDP glucose flavonoid 3-glucosyltransferase (UGT3GT)) participating in both branches (Fig. 2). In *I. nil*, *I. purpurea* and *I. tricolor*, virtually all of the flux normally flows down the cyanidin branch of the pathway, producing blue flowers. However, mutations in these species that block the cyanidin branch produce red pigments and flowers (Fig. 1c), indicating that these downstream enzymes are substrate generalists capable of metabolizing both the hydroxylated and non-hydroxylated intermediates^{16,17}.

Thin-layer chromatography of anthocyanidins from several *Mina* species reveals that the red floral pigments of these species are derived from pelargonidin (Fig. 1b, c). The evolutionary transition to red flowers in this group was thus brought about by inactivation of the cyanidin branch, which can be accomplished in two ways, either by inactivation of F3'H or by the evolution of substrate specificity—loss of the ability to metabolize hydroxylated substrates—by one of the downstream enzymes, DFR or ANS. Substrate specificity of DFR has been demonstrated in several unrelated genera, including *Petunia*¹⁸ and *Arabidopsis*¹⁹. Here we show that both of these types of change have occurred in the lineage leading to *I. quamoclit*.

The single-copy *F3'h* gene from *I. quamoclit* exhibits a high sequence similarity to its homologue from *I. purpurea*. The two sequences exhibit no indels and are 94.7% similar at the nucleotide level and 93.5% similar at the amino-acid level. Although no obvious features (for example, premature stop codons or frame shifts) indicate that the gene is non-functional, our experiments

suggest there has been both regulatory and functional inactivation of $F3'H$ in *I. quamoclit*.

Regulatory inactivation is revealed by slot-blot assays using labelled cDNA from floral bud tissue as a probe. These assays indicate that $F3'h$ expression levels are dramatically reduced in *I. quamoclit* compared with the blue-flowered *I. purpurea*, whereas other genes of the anthocyanin pathway are expressed at comparable levels (Fig. 3d). Reductions of this magnitude in the expression of other anthocyanin genes block pigment production in a variety of *Ipomoea* species²⁰. Absence of cyanidin-based pigments in *I. quamoclit* can be explained entirely by a lack of $F3'H$ in developing flowers and a consequent redirecting of flux down the pelargonidin branch of the pathway.

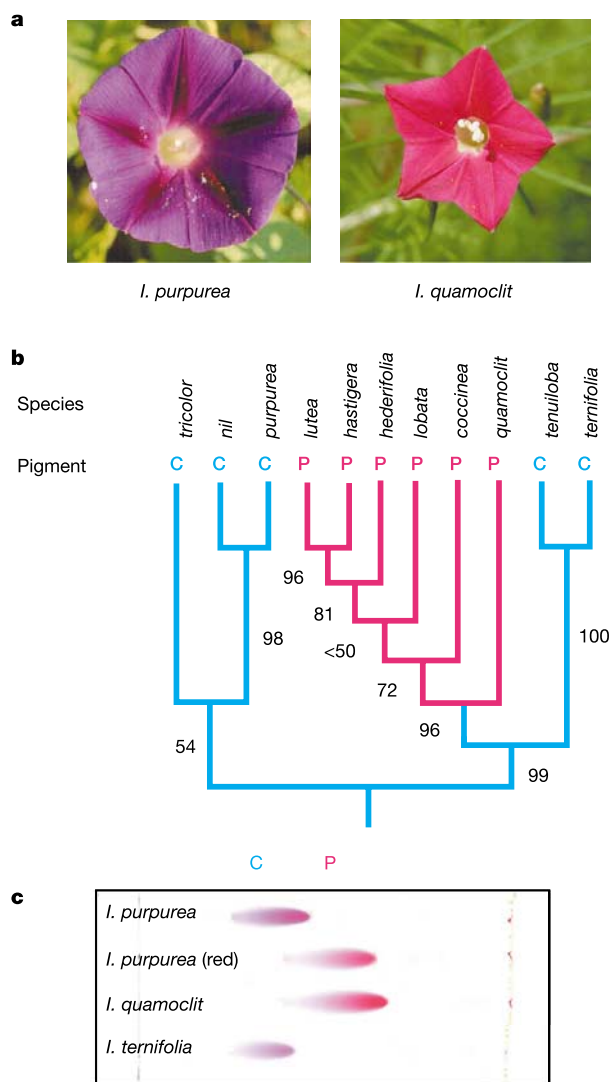


Figure 1 Flower colours and floral pigments of *Ipomoea* species. **a**, *I. purpurea* exhibits blue/purple flowers and inserted stigma and anthers, whereas *I. quamoclit* exhibits red flowers and exerted stigma and anthers. **b**, Phylogenetic relationships among closely related *Ipomoea* species, based on ribosomal internal transcribed sequence. Red branches: species in subgenus *Quamoclit*, section *Mina*, which have red flowers with bird-pollination syndrome. Blue branches: blue/purple flowers and bee-pollination syndrome. Pigment: floral anthocyanin pigments are derived either from cyanidin (C) or pelargonidin (P). Bootstrap support (out of 100 replicates) for the branches is indicated by the numbers next to the branches. **c**, Thin-layer chromatogram illustrating anthocyanin pigments from three species. *I. purpurea* corresponds to the normal blue/purple-flowered morph. *I. purpurea* (red) corresponds to a red-flowered mutant in which $F3'h$ is inactivated^{16,17}.

Functional inactivation of *I. quamoclit* $F3'H$ is suggested by complementation assays, in which we transformed the $F3'h$ -coding region from *I. quamoclit* and *I. purpurea* (this copy serving as a control) linked to a constitutive promoter into an *Arabidopsis thaliana* mutant strain lacking a functional copy of $F3'h$ ¹⁹. Although the *I. purpurea* $F3'h$ gene complemented the *A. thaliana* knockout in four independent transformations, and produced anthocyanins in the leaves of nutrient-stressed plants, no such complementation occurred in any of four transformations with *I. quamoclit* $F3'h$ (Fig. 3c). Moreover, no anthocyanins were detectable, by extraction and thin-layer chromatography, in the leaves of any of the *I. quamoclit* $F3'h$ transformants. Although we did not quantify transcript levels using quantitative polymerase chain reaction (PCR), standard reverse-transcribed PCR (RT-PCR) on leaf RNA from the four transformants of each species yielded bands of comparable intensity on agarose gels, indicating that failure to complement is probably not caused by markedly reduced transcript levels.

This failure to complement does not appear to be due to an inactivation of catalytic activity in the *I. quamoclit* $F3'H$. Qualitative enzyme assays using thin-layer chromatography indicate that the *I. quamoclit* gene is capable of metabolizing its natural substrate dihydrokaempferol (Fig. 3a). Quantitative assays suggest that the efficiency of the *I. quamoclit* $F3'H$ on this substrate may be slightly lower than that of the *I. purpurea* $F3'H$, but this difference is not statistically significant (Fig. 3b). The catalytic activity of the *I. quamoclit* $F3'H$ thus is not substantially impaired. This result suggests that some aspect of the *I. quamoclit* $F3'H$ enzyme other than its catalytic activity (for example, cellular localization) is defective, preventing it from functioning properly *in vivo*. Although we cannot completely rule out the possibility that failure to complement is an experimental artefact of heterologous transformation (for example, due to the lack of a necessary co-factor present in *I. quamoclit* but absent in both *I. purpurea* and *A. thaliana*), we believe this possibility is unlikely because functional anthocyanin pathway genes from very distantly related species (for example, maize) typically complement *A. thaliana* mutants¹⁹.

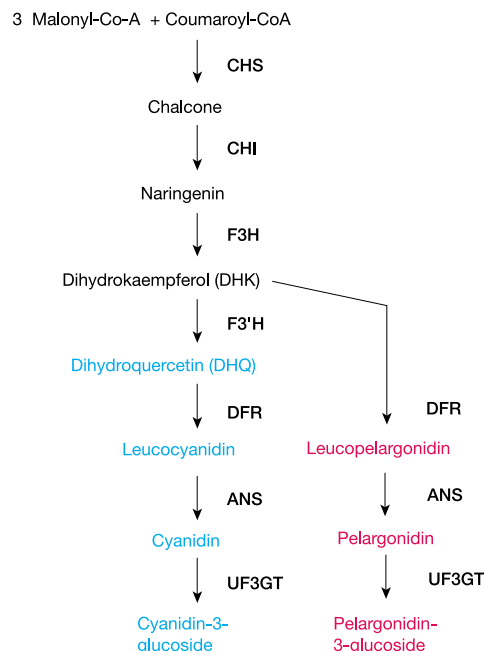


Figure 2 The core anthocyanin biosynthetic pathway. Names in lower case represent pathway intermediates and products. CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavonone 3-hydroxylase. Purple intermediates and products indicate the cyanidin branch of the pathway. Red intermediates and products indicate pelargonidin.

Several properties of DFR demonstrate that evolutionary changes in this enzyme in *I. quamoclit* are sufficient to block the production of cyanidin-based pigments. In *Ipomoea*, *Dfr* is represented by a small family of three paralogues, of which only *Dfr-B* is expressed in floral tissue²¹. The sequence of *Dfr-B* that we cloned from *I. quamoclit* is 95% similar in nucleotide composition and 93% similar in amino-acid composition to its *I. purpurea* orthologue in regions outside of indels. It also exhibits four small in-frame insertions totalling 27 nucleotides, as well as a large 98-nucleotide insertion at the extreme 3' end of the gene (Fig. 4a). This insertion, which is highly similar to the 98 base pairs (bp) just 3' of it, causes a frame shift, which converts a TAA sequence into a stop codon 59 bp upstream of the original stop codon. In addition, two amino-acid substitutions occur in a 13 amino-acid portion of the gene that has been implicated in determining substrate specificity (Fig. 4a)^{22,23}.

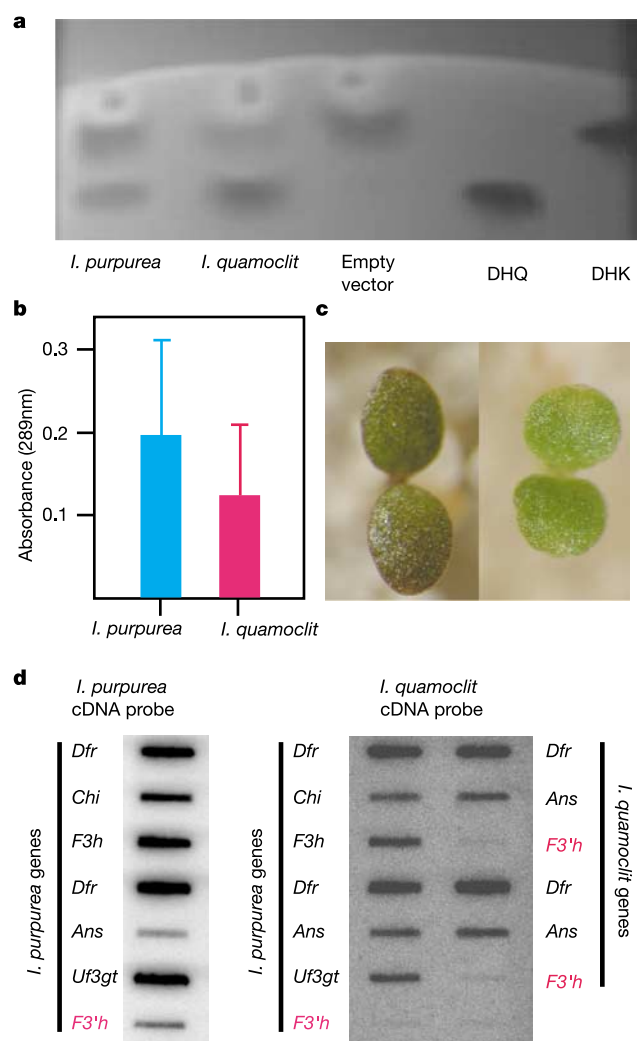


Figure 3 Functional and expression analysis of *F3'h*. **a**, Thin-layer chromatogram illustrating that *F3'h* from both *I. purpurea* and *I. quamoclit* metabolize dihydrokaempferol (DHK) to produce dihydroquercetin (DHQ) in *in vitro* assays. **b**, Mean production of DHQ in *in vitro* enzyme assays. Bars are one standard error. **c**, Accumulation of anthocyanins in *F3'h*-null *A. thaliana* seedlings carrying the transgene *F3'h* from (left) *I. purpurea* (substantial accumulation) and (right) *I. quamoclit* (no visible accumulation). **d**, Anthocyanin gene expression patterns. *F3'h* is highlighted with a red label. Left panel, spotted *I. purpurea* genes probed with *I. purpurea* cDNA. All genes, including *F3'h*, exhibit substantial expression. Right panel, left lane contains spotted genes from *I. purpurea*, right lane genes from *I. quamoclit*. Note that only *Dfr*, *Ans* and *F3'h* have been spotted (each gene twice), because the other anthocyanin genes have not been cloned from *I. quamoclit*. Both lanes of right panel were probed with *I. quamoclit* cDNA.

To determine whether any of these differences influence the ability of the *I. quamoclit* DFR-B to metabolize dihydroquercetin (DHQ) or dihydrokaempferol (DHK), precursors to cyanidin and pelargonidin, respectively, we performed complementation tests and *in vitro* enzyme assays. For complementation, *Dfr-B* from *I. quamoclit* and from *I. purpurea* as a control were transformed into a strain of *A. thaliana* carrying a knockout allele of the sole copy of *Dfr*. In four independent transformations, the *I. purpurea* *Dfr* complemented this knockout and produced pigmented leaves (Fig. 4b). By contrast, in four independent transformations the *I. quamoclit* *Dfr* failed to complement (Fig. 4b), despite substantial expression of the transgene in leaf tissue. Because *A. thaliana* produces only cyanidins (the pelargonidin branch of the pathway is inactive)¹⁹, this failure to complement indicates that the *I. quamoclit* DFR has lost the ability to metabolize DHQ. However, because pelargonidin is produced in the red flowers of *I. quamoclit*, this DFR-B clearly retains the ability to metabolize DHK. Results of *in vitro* enzyme assays support this conclusion. *I. purpurea* DFR metabolizes the two substrates with roughly equal efficiency, whereas the efficiency at which DHQ is metabolized by *I. quamoclit* DFR is markedly reduced, compared with DHK (Fig. 4c). *I. quamoclit* DFR has thus evolved to become a substrate specialist, and this specificity has introduced another block to the cyanidin branch of the anthocyanin pathway in *I. quamoclit*.

Phylogenetic analyses indicate that production of blue/purple,

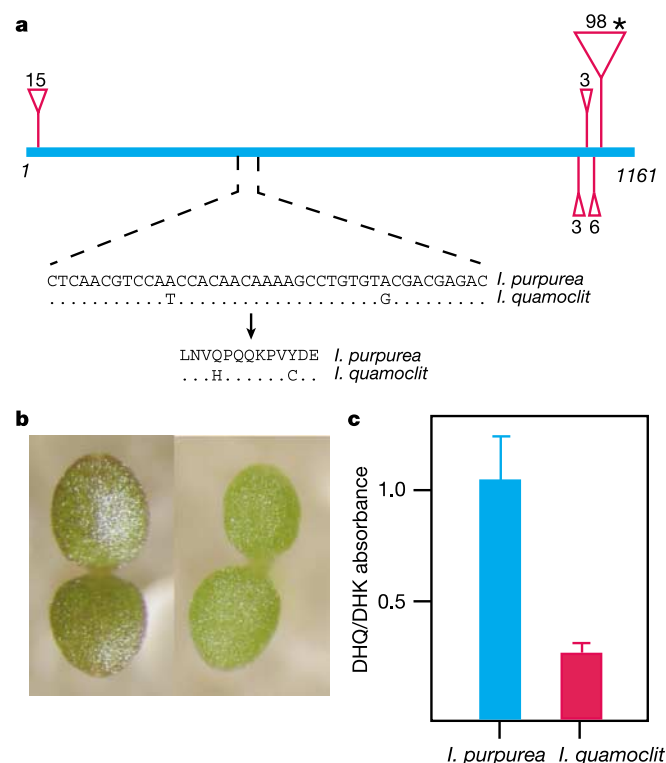


Figure 4 Functional analysis of *Dfr*. **a**, Structural organization. Blue bar, coding region of *I. purpurea* gene. Red open triangles, insertions in *I. quamoclit* (sizes (bp) indicated at bases of triangles). Italic numbers, initial and final bp number. Asterisk, novel stop codon in *I. quamoclit* sequence. Also shown are the nucleotide and amino-acid sequences of the purported substrate specificity-determining region^{22,23}. **b**, Accumulation of anthocyanins in *Dfr*-null *A. thaliana* seedlings carrying the transgene *Dfr* from (left) *I. purpurea* (substantial accumulation) and (right) *I. quamoclit* (no visible accumulation). **c**, Mean and standard error of the ratio of DHQ production to DHK production in *in vitro* assays. The *I. purpurea* enzyme metabolizes DHQ and DHK with equal efficiency, whereas the *I. quamoclit* enzyme metabolizes DHQ with only about quarter the efficiency it metabolizes DHK. The difference in ratio between species is highly significant ($P < 0.01$, *t*-test).

cyanidin-based pigments was the ancestral state in *Ipomoea*^{7,8}, and also in *I. tenuiloba* and *I. ternifolia*, members of the sister clade to the *Mina* section (Fig. 1). The change from blue to red flowers in the ancestral lineage of the *Mina* group, which is associated with a transition from bee to bird pollination, appears to have been brought about by inactivation of the cyanidin branch of the anthocyanin biosynthetic pathway and a redirecting of flux down the pelargonidin branch. We have shown that in *I. quamoclit* this inactivation involves at least two genetic changes: (1) reduction of *F3'h* expression to barely detectable levels; and (2) a marked reduction in the capacity of DFR to metabolize DHQ. In addition, our experiments also suggest that the functionality of the enzyme *F3'H* may be impaired *in vivo*. Because each of these changes is probably sufficient to block cyanidin production, there has evolved in the cyanidin branch at least one (and probably two) redundant block, that is, a block that presumably evolved after the flower colour had changed to red.

Relaxation of selection following the initial blockage of the cyanidin branch thus seems to have facilitated further degeneration in that pathway. This degeneration makes it very unlikely that *I. quamoclit* would be able to re-evolve cyanidin-based blue/purple pigments, since this would require at least two (restoration of *F3'h* expression and either restoration of the ability of DFR to metabolize DHQ or recruitment of expression of a paralogous copy of DFR in flowers), and possibly three (restoration of function to *F3'H*), simultaneous mutations to restore a functional cyanidin branch, and perhaps more if the enzymes ANS and UF3GT have evolved substrate specificity like that seen in DFR. The potential for evolutionary change in an important ecological characteristic seems to have been restricted in *I. quamoclit* by genetic changes that are a consequence of adaptation in that characteristic. □

Methods

Plant materials

Leaf and floral tissue from *I. quamoclit* and *I. purpurea* were collected from populations of these species near Durham, North Carolina.

Pigment identification

Anthocyanidins were extracted from flowers by boiling corollas in 2 M HCl, then extracting with iso-amyl alcohol. This extract was lyophilized, then resuspended in methanol with 1% HCl. Anthocyanidins were separated using thin-layer chromatography on cellulose-coated glass plates in forestal solvent²⁴.

Cloning of *F3'h* and *Dfr-B*

Full-length clones of the entire coding regions of *F3'h* and *Dfr-B* were obtained using RT-PCR and primers based on known sequences from *I. purpurea*. Template cDNA was obtained from floral buds. PCR products were cloned into a TOPO vector and subcloned into the transformation vector *pBI 1.4t* containing the constitutive CaMV 35S promoter and kanamycin and gentamycin resistance for complementation analysis. Sequences of the *I. quamoclit* genes have been deposited in GenBank (accession numbers AY463156 and AY463157).

Complementation analyses

To determine whether *F3'h* and *Dfr* are functional *in vivo*, we employed the complementation analysis using *A. thaliana* strains that were homozygous for a null mutant of the gene to be tested (either *tt7-1* for *F3'h* or *tt3-1* for *Dfr*)¹⁹. Genes in the vector *pBI 1.4t* were transformed first into the *Agrobacterium tumefaciens* strain GV3101 using a freeze-thaw protocol²⁵, then into the appropriate *A. thaliana* mutant strain using the standard dip protocol²⁶. Transformants were selected by growing on MS medium containing kanamycin. Anthocyanin phenotypes of T₂ plants were scored by germinating seeds in sand under high light. DNA was extracted from mature T₂ plants and scored for the presence of the transgene by PCR cloning, followed by sequencing. Expression of the transgene was assayed by RT-PCR using cDNA from seedlings grown in sand under high light.

Enzyme assays

Clones of *Dfr* and *F3'h* from both *I. purpurea* and *I. quamoclit* in the *pBI 1.4t* vector were expressed in *Escherichia coli*. Proteins were extracted from cell culture by disruption in 100–200 µl HEPES buffer with 10 µl of 100 mg ml⁻¹ lysozyme for 30 min on ice. *F3'H* activity was measured using a previously described protocol^{18,27}. Three reactions of each type were run. DFR activity was also assayed using a previously described method²⁸. Seven reactions of each type were run.

F3'h expression assays

To determine the expression levels of *F3'h* and *Dfr*, relative to other anthocyanin structural

genes, DNA slot-blots were constructed using a Bio-Dot SF Microfiltration Apparatus (Bio-Rad) following the manufacturer's instructions. Anthocyanin genes obtained from *I. purpurea* and *I. quamoclit* were spotted onto a Zeta-Probe membrane (Bio-Rad) and fixed by exposure to ultraviolet light. ³²P-labelled cDNA probes were constructed from total RNA extracted from bud tips (the distal half of the bud) of *I. purpurea* and *I. quamoclit* that were 1 day before opening. Membranes were probed overnight at 65 °C and washed according to the manufacturer's instructions. The probe signal was captured and visualized using a phosphorimager (Molecular Dynamics).

Phylogenetic analysis

A phylogeny of 13 *Ipomoea* species (including those species shown in Fig. 1 plus *I. cairica* and *I. sepiaria* as outgroups) was determined based on the sequence of the internal transcribed spacers of nuclear ribosomal DNA (ITS). ITS sequences were obtained from references 9, 10 and 28. A maximum likelihood analysis based on the HKY85+G model was performed in PAUP* version 4.0b10²⁹. Branch support was estimated with 100 bootstrap replicates.

A detailed description of the methods is provided in Supplementary Information.

Received 9 January; accepted 9 March 2004; doi:10.1038/nature02489.

- Grant, V. *Organismic Evolution* (W.H. Freeman, New York, 1977).
- Dobzhansky, T. *Genetics of the Evolutionary Process* (Columbia, New York, 1970).
- Gould, S. J. *The Structure of Evolutionary Theory* (Harvard, Cambridge, MA, 2002).
- Bull, J. J. & Charnov, E. L. On irreversible evolution. *Evolution* **39**, 1149–1155 (1985).
- Gould, S. J. Dollo on Dollo's law: irreversibility and the status of evolutionary laws. *J. Hist. Biol.* **3**, 189–212 (1970).
- Wagner, G. P. The logical structure of irreversible systems transformations: a theorem concerning Dollo's law and chaotic movement. *J. Theor. Biol.* **96**, 337–346 (1982).
- Zufall, R. A. *Evolution of red flowers in Ipomoea*. Thesis, Duke University (2003).
- McDonald, A. Origin and diversity of Mexican Convolvulaceae. *Annales. Inst. Biol. UNAM, Ser. Bot.* **62**, 65–82 (1991).
- Miller, R. E., Rausher, M. D. & Manos, P. S. Phylogenetic systematics of *Ipomoea* (Convolvulaceae) based on ITS and *waxy* sequences. *Syst. Bot.* **24**, 209–227 (1999).
- Miller, R. E., McDonald, J. A. & Manos, P. S. Systematics of *Ipomoea* subgenus *Quamoclit* (Convolvulaceae) based on ITS sequence data and a Bayesian phylogenetic analysis. *Am. J. Botany* (in the press).
- Machado, I. C. S. & Sazima, M. A comparative study in floral biology of two weed species *Ipomoea hederifolia* and *I. quamoclit* Convolvulaceae. *Rev. Brasil. De Biologia* **47**, 425–436 (1987).
- Lava, C. & Ornelas, J. F. Preferential nectar robbing of flowers with long corollas: Experimental studies of two hummingbird species visiting three plant species. *Oecologia* **128**, 263–273 (2001).
- Faegri, D. T. & van der Pijl, L. *The Principles of Pollination Ecology* (Pergamon, Oxford, 1966).
- Saito, N. et al. Acylated cyanidin glycosides in the violet-blue flowers of *Ipomoea purpurea*. *Phytochem.* **40**, 1283–1289 (1995).
- Saito, N. et al. Acylated pelargonidin glycosides in red-purple flowers of *Ipomoea purpurea*. *Phytochem.* **43**, 1365–1370 (1996).
- Hoshino, A. et al. Spontaneous mutations of the flavonoid 3'-hydroxylase gene conferring reddish flowers in the three morning glory species. *Plant Cell Physiol.* **44**, 990–1001 (2003).
- Zufall, R. A. & Rausher, M. D. The genetic basis of a flower-color polymorphism in the common morning glory, *Ipomoea purpurea*. *J. Hered.* **94**, 442–448 (2003).
- Forkmann, G. & Ruhnau, B. Distinct substrate specificity of dihydroflavonols 4-reductase from flowers of *Petunia hybrida*. *Z. Naturforsch.* **42**, 1146–1148 (1987).
- Dong, X., Braun, E. L. & Grotewold, E. Functional conservation of plant secondary metabolic enzymes revealed by complementation of *Arabidopsis* flavonoid mutants with maize genes. *Plant Physiol.* **127**, 46–57 (2001).
- Clegg, M. T. & Durbin, M. L. Tracing floral adaptations from ecology to molecules. *Nature Rev. Genet.* **4**, 206–215 (2003).
- Inagaki, Y. et al. Genomic organization of the genes encoding dihydroflavonol 4-reductase for flower pigmentation in the Japanese and common morning glories. *Gene* **226**, 181–188 (1999).
- Beld, M., Martin, C., Huits, H., Stuitje, A. R. & Gerats, A. G. M. Flavonoid synthesis in *Petunia hybrida*: partial characterization of dihydroflavonols-4-reductase genes. *Plant Mol. Biol.* **13**, 491–502 (1989).
- Johnson, E. T. et al. Alteration of a single amino acid changes the substrate specificity of dihydroflavonols 4-reductase. *Plant J.* **25**, 325–333 (2001).
- Harbourne, J. B. *Comparative Biochemistry of the Flavonoids* (Academic Press, London, 1967).
- An, G., Ebert, P. R., Mitra, A. & Ha, S. B. in *Plant Molecular Biology Manual* (eds Gelvin, S. B. & Schilperoort, R. A.) **A3**, 1–19 (Kluwer, Dordrecht, 1988).
- Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743.
- Stafford, H. A. & Lester, H. H. Enzymatic and nonenzymatic reduction of (+)-dihydroquercetin to its 3,4-diol. *Plant Physiol.* **70**, 695–698 (1982).
- Manos, P. S., Miller, R. E. & Wilkin, P. Phylogenetic analysis of *Ipomoea*, *Argyriaea*, *Stictocardia*, and *Turbina* suggests a generalized model of morphological evolution in morning glories. *System. Bot.* **26**, 585–602 (2001).
- Swofford, D. L. *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Version 4.0 (Sinauer Associates, Sunderland, Massachusetts, 2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This work was supported by NSF grants.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.Z. (bzufall@smith.edu).

The limits to tree height

George W. Koch¹, Stephen C. Sillett², Gregory M. Jennings²
& Stephen D. Davis³

¹Department of Biological Sciences and the Merriam-Powell Center for Environmental Research, Northern Arizona University, Flagstaff, Arizona 86011, USA

²Department of Biological Sciences, Humboldt State University, Arcata, California 95521, USA

³Natural Science Division, Pepperdine University, Malibu, California 90263-4321, USA

Trees grow tall where resources are abundant, stresses are minor, and competition for light places a premium on height growth^{1,2}. The height to which trees can grow and the biophysical determinants of maximum height are poorly understood. Some models predict heights of up to 120 m in the absence of mechanical damage^{3,4}, but there are historical accounts of taller trees⁵. Current hypotheses of height limitation focus on increasing water transport constraints in taller trees and the resulting reductions in leaf photosynthesis⁶. We studied redwoods (*Sequoia sempervirens*), including the tallest known tree on Earth (112.7 m), in wet temperate forests of northern California. Our regression analyses of height gradients in leaf functional characteristics estimate a maximum tree height of 122–130 m barring mechanical damage, similar to the tallest recorded trees of the past. As trees grow taller, increasing leaf water stress due to gravity and path length resistance may ultimately limit leaf expansion and photosynthesis for further height growth, even with ample soil moisture.

According to the cohesion-tension theory, water transport in plants occurs along a gradient of negative pressure (tension) in the dead, tube-like cells of the xylem, with transpiration, water adhesion to cell walls, and surface tension providing the forces necessary to lift water against gravity⁷. Height growth may slow if the xylem tension and therefore leaf water potential (Ψ) predicted for great heights, ≤ -2 MPa (ref. 7), reduces sufficiently the positive pressure (turgor) necessary for expansion of living cells or increases the risk of xylem cavitation—cavitation is the formation of embolisms that reduce hydraulic conductivity and can cause branch dieback and plant death^{8,9}. Many trees respond to Ψ below -1 MPa by decreasing the aperture of microscopic pores (stomata) in leaves through which water vapour is lost in transpiration and carbon dioxide (CO_2) is gained in photosynthesis¹⁰. Reduced stomatal conductance can decrease cavitation risk and turgor loss, but it also limits photosynthesis. Thus, as trees grow taller, maintenance of favourable water status might progressively slow height growth by reducing photosynthetic carbon gain^{4,6}.

We accessed the crowns of redwoods to measure water stress and photosynthesis and to collect samples for laboratory analyses. Within individual trees, the xylem pressure of small, foliated branches measured during the dry season (late September to early October) was strongly correlated with height (Fig. 1a). The gradient before dawn, when transpiration was negligible, averaged $-0.0096 \pm 0.0007 \text{ MPa m}^{-1}$ for five trees over 110 m tall ($R^2 > 0.97$, $P < 0.0001$), nearly identical to the hydrostatic gradient due to gravity ($-0.0098 \text{ MPa m}^{-1}$) as predicted by the cohesion-tension theory⁷. The slope of the xylem pressure–height relationship was slightly steeper ($-0.0106 \pm 0.0022 \text{ MPa m}^{-1}$) at midday when the evaporative gradient and transpiration were high. The minimum xylem pressure (that is, maximum tension) recorded in the highest branches sampled ($108 \pm 1.2 \text{ m}$) averaged $-1.84 \pm 0.04 \text{ MPa}$. The importance of height *per se* for water potential was evident in that nearly two-thirds of the midday xylem pressure was due to gravity.

Reduced water potential due to soil drought causes a decline in the turgor of living plant cells that is necessary for cell growth and leaf expansion¹¹. To determine if this also occurs as water potential declines with height, we estimated turgor at dry-season water potentials from pressure–volume measurements. Turgor (in MPa) declined linearly with height, h , as $\text{turgor} = -(0.0074 \pm 0.0004)h + (1.30 \pm 0.07)$, $n = 4$ trees, ranging from 0.93 MPa at 50 m to 0.48 MPa at 110 m. At night when xylem pressure increased, the turgor gradient was less steep, $\text{turgor} = -(0.0044 \pm 0.0023)h + (1.39 \pm 0.19)$, and turgor was 0.3–0.4 MPa higher than at midday.

Given the role of turgor in leaf expansion, its reduction with height may underlie the distinct vertical gradient in leaf structure in redwoods (Fig. 2). Leaf shape varied from large and expanded in the lower crown to small and scale-like at the treetop. We quantified this variation in terms of the leaf mass:area ratio (LMA, g m^{-2}), which increased exponentially over a fourfold range with height (Fig. 1b, $\text{LMA} = (37.1 \pm 12.3)\exp(0.0260 \pm 0.0030)h$, $0.88 \leq R^2 \leq 0.99$, $0.0001 \leq P \leq 0.003$, $n = 5$ trees). At 112 m, LMA was similar to the highest published value for terrestrial plants¹². Height-related variation in LMA has been attributed to light level in forest canopies^{13,14}. In our study trees, we found that the direct site factor (DSF), an index of direct solar radiation based on hemispherical photographs, decreased by 14% of the value at 110 m for a 10-m decrease in height. Relative to water potential, the influence of light on LMA was small, however; DSF added only 4% to the explained variation of within-crown LMA in a multiple regression analysis including DSF ($P = 0.0025$) and predawn xylem pressure ($P < 0.0001$) as independent variables (adjusted $R^2 = 0.88$, $n = 33$ samples from five redwoods over 110 m tall). The following observations (Fig. 3) also support the hypothesis that water relations are more important than light environment in determining leaf structure in redwood: (1) leaves of a 2-m-tall epiphytic redwood rooted in soil near the top of a 95-m-tall redwood were much more

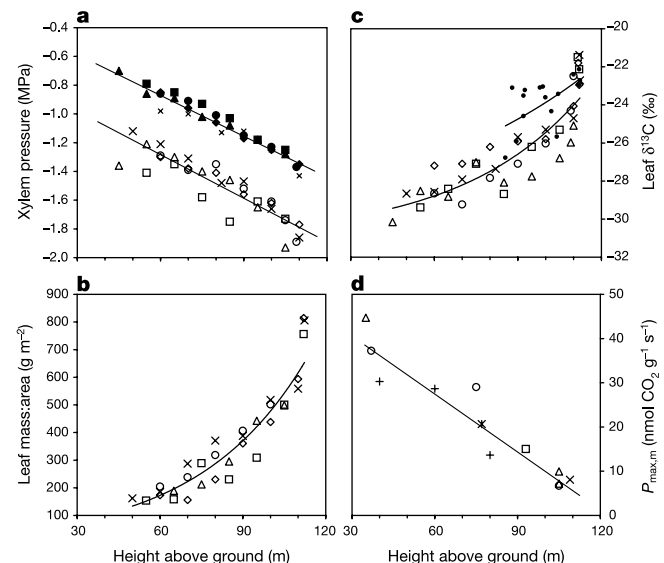


Figure 1 Variation with height in physiological and structural features of redwood trees at Humboldt Redwoods State Park, California. **a**, Xylem pressure of small branches measured at predawn (upper group) and midday (lower group) during September and October 2000. The upper line is the expected gravitational pressure gradient with the same y-intercept as the average of the 5 trees. **b**, Leaf mass:area ratio (g m^{-2}) of second-year internodes increases with height. **c**, Foliar carbon isotope composition ($\delta^{13}\text{C}$, ‰) increases with height within the crowns of 5 trees over 110 m tall and among the tops (filled circles) of 16 trees from 85 to 113 m tall. **d**, Light-saturated photosynthetic rate per unit mass ($\mu\text{mol CO}_2 \text{ g}^{-1} \text{ s}^{-1}$) decreases with height. The regression line is fitted to data from six trees. Different symbol types denote different trees and are consistent for **a–d**.

expanded than leaves of the host tree in the same light environment, and (2) when a fallen branch from the upper crown of a tall redwood was potted in wet soil and allowed to root with the branch exposed to high light, the new leaves produced were much more expanded than the existing leaves. It is likely that in very tall trees, components of water potential, notably turgor, are important determinants of LMA and related anatomical features, just as they are along soil moisture gradients¹².

To assess further the physiological consequences of declining water potential with height, we measured stable carbon isotope composition ($\delta^{13}\text{C}$) and leaf photosynthesis. The $\delta^{13}\text{C}$ value (‰) of plant tissues expresses the photosynthetic discrimination against $^{13}\text{CO}_2$ compared with $^{12}\text{CO}_2$ and is a common metric of long-term water stress, ranging from -20‰ to -34‰ in plants such as redwood that have the C_3 photosynthetic pathway¹⁵. When water stress reduces stomatal conductance, CO_2 concentration in the intercellular air spaces of leaves (C_i) declines, and this lessens the enzymatic discrimination against $^{13}\text{CO}_2$, causing $\delta^{13}\text{C}$ to increase. In redwood, foliar $\delta^{13}\text{C}$ (‰) was correlated with height (Fig. 1c, exponential fits of the form $\delta^{13}\text{C} + 31 = b_1 \exp(b_0 h)$, $b_1 = 0.681 \pm 0.434$, $b_0 = 0.033 \pm 0.005$, $0.78 \leq R^2 \leq 0.94$, $0.0001 \leq P \leq 0.002$). The highest $\delta^{13}\text{C}$ was always observed at the treetop and averaged $-22.2 \pm 0.6\text{‰}$ at $\geq 110\text{ m}$. This is the highest published foliar $\delta^{13}\text{C}$ for tall trees and is close to the apparent limit for C_3 plants^{16–18}. We also found a significant relationship of $\delta^{13}\text{C}$ (‰) to height for the treetop foliage of 13 trees from 85 m to 110 m tall plus 3 trees taller than 112 m (Fig. 1c, $\delta^{13}\text{C} + 31 = 1.928 \exp(0.0131 h)$, $R^2 = 0.321$, $P = 0.022$, $n = 16$). That $\delta^{13}\text{C}$ was higher (less negative) at the tops of trees than at the same height within tree crowns probably indicates the effect of shading in reducing foliar $\delta^{13}\text{C}$ (ref. 15). Nonetheless, patterns both within and among trees demonstrate a strong increase of $\delta^{13}\text{C}$ with height in redwood, indicative of increasing water stress.

High values of $\delta^{13}\text{C}$ occur when stomatal conductance is strongly limiting to photosynthesis, as in plants experiencing low water potentials due to soil moisture stress^{15,16}. Rather than soil moisture stress, however, it is likely that in tall trees the reduction in water potential due to gravity and path length resistance causes stomatal conductance to increasingly limit photosynthesis with height^{6,19}. The treetop $\delta^{13}\text{C}$ of about -22‰ corresponds to a flux-weighted C_i of $\sim 160\text{ p.p.m.}$ during the assimilation of CO_2 into new biomass. This is similar to the average daily C_i estimated from our *in situ* gas exchange measurements near the tops of two 112-m-tall trees during autumn ($170 \pm 11\text{ p.p.m.}$). Thus, integrated $\delta^{13}\text{C}$ and

instantaneous gas exchange indicate that stomatal conductance is increasingly limiting to photosynthesis with height, as reported for other conifers up to 65 m tall^{19,20}.

Laboratory gas exchange measurements of foliage cut from different heights and re-hydrated to uniformly high water potentials enabled us to examine the consequences of height for photosynthesis in the absence of a direct influence of low water potential. Light-saturated photosynthesis per unit leaf area ($P_{\text{max,a}}$) did not vary with height ($R^2 = 0.012$, $P = 0.78$), averaging $5.6\text{ }\mu\text{mol CO}_2\text{ m}^{-2}\text{ s}^{-1}$. Photosynthesis per unit leaf mass ($P_{\text{max,m}}$) decreased with height (Fig. 1d, $P_{\text{max,m}} = -0.455h + 55.3$, $R^2 = 0.88$, $P = 0.0002$), however, indicating a lower potential photosynthetic return on biomass invested in leaves at greater heights. The $P_{\text{max,m}}$ of 110-m foliage was 28% of that at 80 m and only 16% of that at 50 m. Because light levels decline exponentially with depth in forest canopies², actual differences in photosynthesis per unit biomass are probably smaller than indicated by our

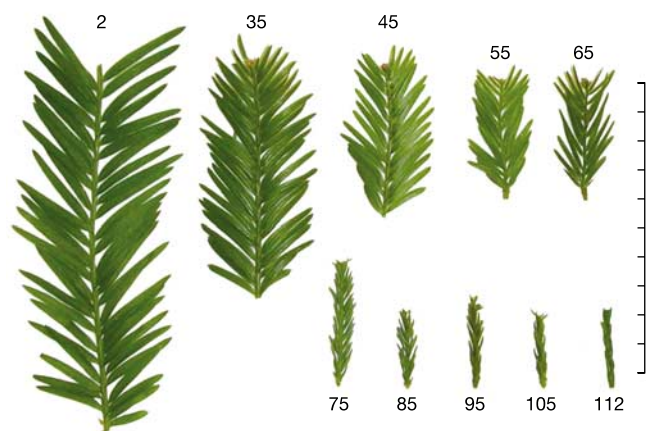


Figure 2 Variation in leaf structure with height in redwood. Leaf length and the angle between the long axis of the leaf and supporting stem segment both decrease with height. Numbers denote the sample height in m. Scale divisions are cm.



Figure 3 Leaf structure can vary independently of light environment. The upper panel shows foliage of an epiphytic redwood (expanded, light-green leaves) and adjacent foliage of the host redwood at 95 m in the same light environment. The lower panel shows the new, more expanded foliage (at branch tips) that developed next to the existing unexpanded foliage on a detached upper crown (>90 m) branch after it was rooted in wet soil and kept in high light. Both examples support the view that variation in light environment explains little of the variation in leaf structure in redwood (see text).

measurements at light saturation. Nonetheless, photosynthetic capacity per unit foliar mass declines with height, and we hypothesize that this results from the observed changes in leaf structure. High LMA is associated with high tissue density and increased allocation of biomass to structure, including thicker cell walls^{12,21}. These changes can increase internal resistance to CO₂ diffusion within leaves^{22,23}, reducing photosynthesis and contributing to high $\delta^{13}\text{C}$ values. We speculate that the universal influence of gravity on water potential gradients in tall trees underlies structural changes in photosynthetic tissues that, along with increased stomatal regulation, reduce photosynthesis and carbohydrate availability for height growth.

Our studies of the tallest redwoods reveal gradients in physiological and structural features that support hypotheses of height limitation due to hydraulic constraints^{4,6}. Water potential, turgor, leaf structure, carbon isotope composition, and photosynthesis all change with height as they do along gradients of soil moisture stress, consistent with a general role for water availability in determining leaf functional traits^{12,21}. Height gradients of these variables also allowed us to address the potential biophysical limit to height in redwood. To estimate a maximum height in the current environment barring mechanical damage, we calculated the height at which the functional variables we examined would reach a limit value (Table 1).

Low water potential affects growth severely when the formation of embolisms by cavitation reduces hydraulic conductivity⁸. Whereas it is not uncommon for shorter trees to operate at water potentials that cause considerable loss of hydraulic conductivity, cavitation avoidance may be critical for height growth in tall trees. Great height may prevent recovery of lost hydraulic function by embolism dissolution, the standard model for which requires that xylem pressures rise to within a few tenths of 0 MPa (ref. 24), higher than is possible in a water column held by tension above a few tens of metres. Our measurements of xylem vulnerability to cavitation in upper crown branches (109 m) of five trees over 110 m tall indicate that loss of hydraulic conductivity begins as xylem pressure drops to -1.9 MPa (Supplementary Information), slightly lower than the lowest pressure we recorded at the tops of the tallest trees at midday. The xylem pressure–height relationship (Fig. 1a, Table 1) estimates a pressure of -1.9 MPa at 122 m, increasing to 132 m for a limit value of -2.0 MPa. It is likely that lower water potentials and cavitation do occur in tall redwoods, mature individuals of which may experience severe droughts during life spans of up to 2,200 years (refs 25, 26). It may be during such episodes that the upper crown dies back, as evidenced by our observations that nearly all very tall redwoods have multiple tops, the original leader having died and been replaced repeatedly.

Surveys of several hundred terrestrial plant species across diverse biomes report values of LMA from 20 g m^{-2} for thin planar leaves of herbaceous species to a maximum of 833 g m^{-2} for the scale-like leaves in *Juniperus monosperma*, a short coniferous tree of arid regions and, like redwood, a member of the Cupressaceae^{12,21}. Using 833 g m^{-2} as the maximum possible for redwood, the LMA–height relationship (Fig. 1b, Table 1) estimates a maximum height of 122 m. If the limiting value of LMA is allowed to increase by 10%, the maximum height increases to 126 m.

For C₃ plants, the apparent limit of foliar $\delta^{13}\text{C}$ is the approxi-

mately -20‰ reported for plants of arid environments^{16–18}. The overall within-crown $\delta^{13}\text{C}$ –height relationship (Fig. 1c, Table 1) estimates a $\delta^{13}\text{C}$ of -20‰ at 130 m. The sensitivity of this estimate to the limiting $\delta^{13}\text{C}$ value is low; heights of 134 m and 125 m are estimated for $\delta^{13}\text{C}$ of -19‰ and -21‰ , respectively.

A linear extrapolation of maximum photosynthesis versus height (Fig. 1d, Table 1) predicts that $P_{\text{max,m}}$ in saturating light would decline to 0 at 125 m. Carbon import from elsewhere in the tree may support early growth of new leaves, but the trend in foliar $\delta^{13}\text{C}$ with height (Fig. 1c) indicates at most a minor quantitative significance of carbon subsidies from the lower crown.

Taken together, these height trends in ecophysiological variables indicate that the maximum height of redwood at our study site in current environmental conditions is 122 to 130 m. The reduction in water potential with height reduces leaf expansion and photosynthesis, the latter directly via increased stomatal regulation, as evidenced by $\delta^{13}\text{C}$, and indirectly by altering leaf structure (LMA), which in turn may further constrain carbon balance. Several additional lines of evidence support a limit to tree height for redwood that is taller than today's tallest trees and near the estimates from our regression analyses. First, our measurements indicate that the tallest redwoods are growing by up to 0.25 m yr^{-1} . Second, over 95% of the original old-growth redwood forest has been logged²⁶, and it is likely that redwoods taller than today's giants were felled⁵. Third, we analysed the height gradient in foliar $\delta^{13}\text{C}$ reported for Douglas-fir²⁰ ($\delta^{13}\text{C} = 0.060h - 27.5$, $R^2 = 0.999$, $P = 0.005$) and estimated by linear extrapolation that $\delta^{13}\text{C}$ would reach -20‰ at 125 m. Finally, the maximum tree height we predict for redwood and Douglas-fir is similar to the 126 m of the tallest reliably measured gymnosperm of the past, a Douglas-fir⁵.

The tallest redwoods today stand in large reserves where intact forest structure sustains moist conditions and buffers trees against wind. The trees in this study, which include the first, second, fourth, sixth and eighth tallest known individuals on Earth, all occur within the largest contiguous old-growth redwood forest remaining (Humboldt Redwoods State Park, California), a reserve protecting 89 of the 116 tallest redwoods. (Measurements of redwoods throughout the species' range in California have found 116 individuals over 107 m; C. Atkins and M. Taylor, personal communication.) At reserves further north and closer to the coast, stronger storms may explain the lower heights (~ 100 m), yet similar relationships of water potential and $\delta^{13}\text{C}$ to height²⁷ as we observed in the tallest redwoods. At the drier inland margin of redwood's natural distribution in northern California, maximum tree height is lower (~ 80 m), yet treetop values of minimum water potential (-1.9 MPa) and maximum $\delta^{13}\text{C}$ (-22‰) are similar to those at 110 m in the tallest redwoods. Thus a similar physiological ceiling may be reached at different physical heights depending on water availability, with storm damage reducing realized heights at sites that are otherwise optimal. Tree height should also vary over time as climate fluctuates, and linking top dieback dates and growth rates to past climate may reinforce our physiological interpretation of height limitation. Climate and atmospheric change will affect the height to which redwoods grow, the outcome depending on the combined effects of elevated atmospheric CO₂ concentration and altered temperature and moisture on tree water relations and carbon balance²⁸. □

Table 1 Maximum height predictions for redwood, *Sequoia sempervirens*

Dependent variable	Equation	Limit value of dependent variable	Maximum height (m)
Ψ , midday (MPa)	$\Psi = -0.00973h - 0.712$	-1.9	122
LMA (g m^{-2})	$\text{LMA} = 37.43\text{exp}(0.0255h)$	833	122
$\delta^{13}\text{C}$ (‰)	$\delta^{13}\text{C} = 0.559\text{exp}(0.0229h) - 31$	-20	130
$P_{\text{max,m}}$ ($\text{nmol g}^{-1}\text{s}^{-1}$)	$P_{\text{max,m}} = -0.434h + 54.3$	0	125

The relationship of physiological and structural variables to height in redwoods at Humboldt Redwoods State Park, California. The equations describe the relationship of the dependent variable to height for data from all study trees combined. See text for explanations of limit values for dependent variables.

Methods

Tree access

We accessed tree crowns by shooting arrows trailing filament over branches with a powerful bow. Rope was then hauled over the branches and climbed via mechanical ascenders. Access to the treetop was achieved by arborist-style techniques. Heights were measured by lowering weighted fibreglass measuring tapes from the treetop to average ground level.

Physiological measurements

Water potential of small branches (≤ 15 cm length) located within 1 to 3 m of the main trunk was measured using a pressure chamber (PMS Instruments). Measurements of photosynthesis used a portable photosynthesis system (LI6400, LiCor) with a 2 cm $\times$ 3 cm chamber with red/blue LED light source. Photosynthesis was measured under controlled conditions: air temperature, $22 \pm 1^\circ\text{C}$; CO_2 concentration, 365 ± 10 p.p.m.; vapour pressure deficit, 1.2 ± 0.2 kPa; light, $\geq 1,400$ $\mu\text{mol photons m}^{-2} \text{s}^{-1}$. Samples used for laboratory measurements of photosynthesis and pressure–volume relationships were cut from different heights, then re-cut immediately under water, allowed to re-hydrate overnight, and then measured. This produced high water potentials (-0.6 ± 0.3 MPa) and allowed comparisons of photosynthetic capacity without the influence of height-related variation in water potential. During these measurements, the C_i values did not differ significantly in foliage from different heights (239 ± 16 p.p.m., $P = 0.42$). Turgor was estimated by the pressure–volume method²⁹.

Morphological measurements

To determine LMA, projected surface areas of 10 second-year internodes from each sample height were measured using a digital surface-area meter (Delta T Instruments). Samples were oven-dried at 70°C , weighed, and mean LMA calculated as g m^{-2} . Area and mass measurements included the entire foliated internode.

Stable carbon isotope composition

$\delta^{13}\text{C}$ of foliage samples was analysed at the Colorado Plateau Stable Isotope Laboratory (<http://www4.nau.edu/cpsil/>). In 2000, second-year internodes were collected at different heights, dried (70°C), ground to 40 mesh, and then a subsample was pulverized, encapsulated in tin, and combusted (CE Instruments NC2100) at $1,000^\circ\text{C}$. The resultant CO_2 was purified and its $^{13}\text{CO}_2/^{12}\text{CO}_2$ ratio was analysed by isotope-ratio mass spectrometry (Delta Plus XL, ThermoQuest Finnigan) in continuous-flow mode. The $\delta^{13}\text{C}$ values were expressed as the relative abundance of ^{13}C versus ^{12}C compared with a standard (Pee Dee Belemnite): $\delta^{13}\text{C} = (R_{\text{sam}}/R_{\text{std}} - 1)1,000\text{‰}$, where R_{sam} and R_{std} are the $^{13}\text{C}/^{12}\text{C}$ ratios in sample and standard, respectively. The standard deviation of repeated measurements of secondary standard material was $<0.1\text{‰}$ (external precision).

Light environment

Hemispherical photographs were taken directly above leaf sample locations throughout tree crowns using a digital camera on a self-levelling mount. Photographs were analysed with WinSCANOPY (v.2002a, Régent Instruments Inc.) to calculate direct site factor, which is the average proportion of direct radiation received during the 12-month growing season.

Received 7 November 2003; accepted 16 February 2004; doi:10.1038/nature02417.

- King, D. A. The adaptive significance of tree height. *Am. Nat.* **135**, 809–828 (1991).
- Waring, R. H. & Schlesinger, W. H. *Forest Ecosystems* (Academic, Orlando, 1985).
- West, G. B., Brown, J. H. & Enquist, B. J. A general model for the structure and allometry of plant vascular systems. *Nature* **400**, 664–667 (1999).
- Friend, A. D. in *Vegetation Dynamics and Global Change* (eds Solomon, A. M. & Shugart, H. H.) 101–115 (Chapman and Hall, New York, 1993).
- Carder, A. C. *Forest Giants of the World, Past and Present* (Fitzhenry & Whiteside, Markham, Ontario, 1995).
- Ryan, M. J. & Yoder, B. J. Hydraulic limits to tree height and tree growth. *Bioscience* **47**, 235–242 (1997).
- Zimmermann, M. H. *Xylem Structure and the Ascent of Sap* (Springer, New York, 1983).
- Tyree, M. T. & Sperry, J. S. The vulnerability of xylem to cavitation and embolism. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **40**, 19–38 (1989).
- Davis, S. D. et al. Shoot dieback during prolonged drought in *Ceanothus* (Rhamnaceae) chaparral of California: a possible case of hydraulic failure. *Am. J. Bot.* **89**, 820–828 (2002).
- Kramer, P. J. & Boyer, J. S. *Water Relations of Plants and Soils* (Academic, San Diego, 1995).
- Taiz, L. & Zeiger, E. *Plant Physiology*, 3rd edn (Sinauer Associates, Sunderland, Massachusetts, 2002).
- Reich, P. B. et al. Generality of leaf trait relationships: a test across six biomes. *Ecology* **80**, 1955–1969 (1999).
- Niinemets, U., Kull, O. & Tenhunen, J. D. An analysis of light effects on foliar morphology, physiology and light interception in temperate deciduous woody species of contrasting shade tolerance. *Tree Physiol.* **18**, 681–696 (1998).
- Bond, B. J., Farnsworth, B. T., Coulombe, R. A. & Winner, W. E. Foliage physiology and biochemistry in response to light gradients in conifers with varying shade tolerance. *Oecologia* **120**, 183–192 (1999).
- Farquhar, G. D., Ehleringer, J. R. & Hubick, K. T. Carbon isotope discrimination and photosynthesis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **40**, 503–537 (1989).
- Ehleringer, J. R. in *Stable Isotopes in Plant Carbon–Water Relations* (eds Ehleringer, J. R., Hall, A. E. & Farquhar, G. D.) 155–172 (Academic, San Diego, 1993).
- Vogel, J. C. in *Stable Isotopes in Plant Carbon–Water Relations* (eds Ehleringer, J. R., Hall, A. E. & Farquhar, G. D.) 29–46 (Academic, San Diego, 1993).
- Van de Water, P. K., Leavitt, S. W. & Betancourt, J. L. Leaf $\delta^{13}\text{C}$ variability with elevation, slope aspect, and precipitation in the southwest United States. *Oecologia* **132**, 332–343 (2002).
- Yoder, B. J., Ryan, M. G., Waring, R. H., Schoettle, A. W. & Kaufmann, M. R. Evidence of reduced

- photosynthetic rates in old trees. *Forest Sci.* **40**, 513–526 (1994).
- McDowell, N. G., Phillips, N., Lunch, C., Bond, B. J. & Ryan, M. G. An investigation of hydraulic limitation and compensation in large, old Douglas-fir trees. *Tree Physiol.* **22**, 763–772 (2002).
- Niinemets, U. Components of leaf dry mass per area—thickness and density—alter leaf photosynthetic capacity in reverse directions in woody plants. *New Phytol.* **144**, 35–47 (1999).
- Parkhurst, D. F. Diffusion of CO_2 and other gases inside leaves. *New Phytol.* **126**, 449–479 (1994).
- Warren, C. R. et al. Transfer conductance in second growth Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) canopies. *Plant Cell Environ.* **26**, 1215–1227 (2003).
- Hacke, U. G. & Sperry, J. S. Limits to xylem refilling under negative pressure in *Laurus nobilis* and *Acer negundo*. *Plant Cell Environ.* **26**, 303–311 (2003).
- Stine, S. Extreme and persistent drought in California and Patagonia during mediaeval time. *Nature* **369**, 546–549 (1994).
26. Noss, R. F. (ed.) *The Redwood Forest: History, Ecology and Conservation of Coast Redwoods* (Island, Washington DC, 2000).
- Jennings, G. M. *Vertical Hydraulic Gradients and the Cause of Foliar Variation in Tall Redwood Trees* Thesis, Humboldt State Univ., Arcata, California (2003).
- Medlyn, B. E. et al. Stomatal conductance of forest species after long-term exposure to elevated CO_2 concentration: a synthesis. *New Phytol.* **149**, 247–264 (2001).
- Boyer, J. S. *Measuring the Water Status of Plants and Soils* (Academic, San Diego, 1995).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This work was supported by the Global Forest Society, the Save-the-Redwoods League, and Northern Arizona University's Organized Research, and permitted by Redwood State and National Parks. J. Amthor, S. Burgess, T. Dawson, A. Fredeen, B. Hungate and H. Mooney provided comments that improved the paper.

Authors' contributions G.K., S. S. and G.J. conceived and conducted the experiments, and G.K. and S.S. analysed the data and co-wrote the paper. S. D. and G. K. conducted the xylem cavitation experiments.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.W.K. (george.koch@nau.edu).

Perceived luminance depends on temporal context

David M. Eagleman^{1,2}, John E. Jacobson^{2,3} & Terrence J. Sejnowski^{2,4}

¹Department of Neurobiology and Anatomy, University of Texas, Houston Medical School, 6431 Fannin Street, Suite 7.046, Houston, Texas 77030, USA

²Howard Hughes Medical Institute at the Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA

³Department of Philosophy and ⁴Division of Biological Sciences, University of California at San Diego, La Jolla, California 92093, USA

Brightness—the perception of an object's luminance—arises from complex and poorly understood interactions at several levels of processing¹. It is well known that the brightness of an object depends on its spatial context², which can include perceptual organization³, scene interpretation⁴, three-dimensional interpretation⁵, shadows⁶, and other high-level percepts. Here we present a new class of illusion in which temporal relations with spatially neighbouring objects can modulate a target object's brightness. When compared with a nearby patch of constant luminance, a brief flash appears brighter with increasing onset asynchrony. Simultaneous contrast, retinal effects, masking, apparent motion and attentional effects cannot account for this illusory enhancement of brightness. This temporal context effect indicates that two parallel streams—one adapting and one non-adapting—encode brightness in the visual cortex.

We report here a novel illusion in which temporal relationships affect brightness perception. Two flashes appeared on either side of a fixation point: one was brief (56 ms), the other long (278 ms; Fig. 1a). Observers reported which flash appeared brighter. When flashes of identical luminance had simultaneous onset, subjects

reported that the brief flash looked dimmer than the long flash (Fig. 1b). This was expected from the Broca–Sulzer effect, in which a flash of brief duration looks dimmer than a physically identical flash presented for a longer duration⁷. However, when the two flashes had simultaneous offset, the brief flash appeared brighter (Fig. 1b; $t = -5.32$, $P = 0.0009$). This surprising brightness enhancement averaged 30% and was seen by all observers tested. When the brief flash occurred at intermediate times (between onset and offset of the long flash), the brightness grew monotonically (Fig. 1c). We call this illusion the temporal context effect (TCE; see the online demonstration at <http://nba.uth.tmc.edu/homepage/eagleman/TCE>).

It could be that the brief flash looks the same in all conditions but is reported as brighter because it is compared with a dimming long flash. To address this possibility we tested whether the long flash dims perceptually. Observers were asked whether they detected a change in flashes that either physically ramped up or down in luminance or stayed constant. Observers reported that a physically

unchanging flash was perceived as unchanging; that is, the observers (at least in these conditions) were adaptation-blind (Fig. 2a). This suggests that relative timing actually changes the brightness of the brief flash. To test this, we presented the stimulus shown in Fig. 2b: a brief flash is onset-matched with two long flashes; a second brief flash is offset-matched with the long flashes. Observers directly compared the two brief flashes against each other. The offset-matched brief flash appeared brighter than the onset-matched brief flash. This effect depends on the presence of the long flashes: in a control experiment, in which only the brief flashes were presented, no brightness difference was seen between them (see Supplementary Information). In a second control, we found that in static displays of three patches, dimming the patches corresponding to the long ones in Fig. 2b did not induce simultaneous contrast effects that could account for the TCE (see Supplementary Information). Thus, the appearance of the brief flash itself changed, as a result of temporal context.

The TCE does not arise from low-level interactions in the retina or lateral geniculate nucleus, because presenting the brief flash to one eye and the long flash to the other produces the full effect (see Supplementary Information). This indicates that the comparison of luminance ratios takes place in primary visual cortex or later, where binocular information first converges. Further, the TCE appears unrelated to apparent motion or meta-contrast masking, because it is surprisingly insensitive to spatial distance (Fig. 2c). This insensitivity can easily be demonstrated by presenting the long and brief

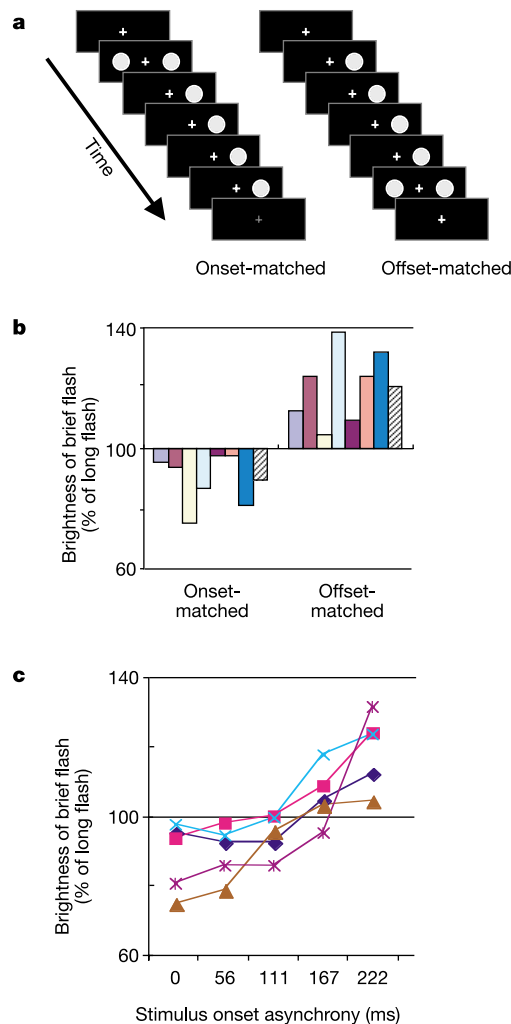


Figure 1 The temporal context effect. **a**, Time slices (56 ms) of two stimulus conditions. **b**, In a two-alternative forced-choice task, observers reported whether the brief flash was brighter or dimmer than the long flash. Brightness on the ordinate is determined by the point of perceived equivalence on a psychometric function, derived from the method of constant stimuli. The long flash had a luminance of 5.91 cd m^{-2} , the brief flash ranged from 3.55 to 8.27 cd m^{-2} and the background was less than 1 cd m^{-2} . $n = 7$. The average is shown with a hatched bar. **c**, Brightness of the brief flash at intermediate stimulus onset asynchronies.

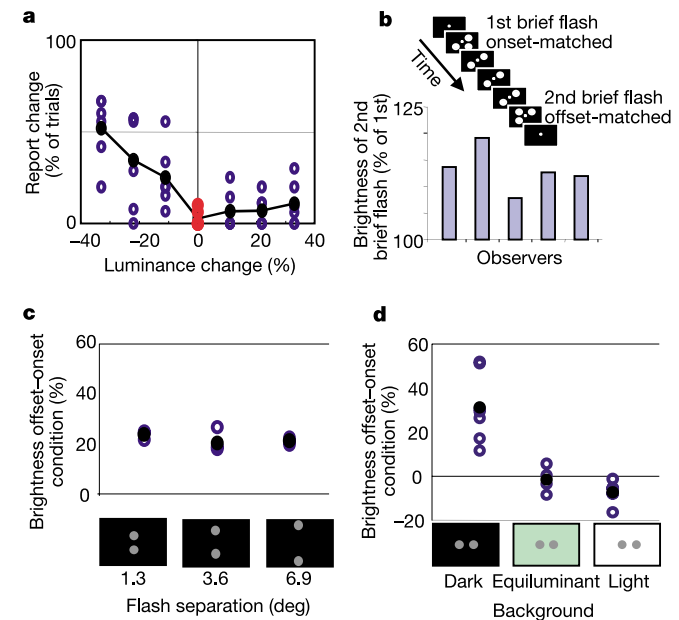


Figure 2 Only the brief flash changes perceptually. **a**, Adaptation-blindness. The abscissa shows the total change of flash luminance during the 278-ms duration (linear ramping). Subjects reported whether the flash did or did not change in brightness during its appearance. Subjects are better at detecting decrements than increments¹⁵, but the emphasis for this study is on the zero point of the x-axis: a physically unchanging flash is perceived as unchanging. Stimulus properties and starting luminance are identical to those of the long flashes in Fig. 1. $n = 6$. **b**, Observers compared the brightness of the diagonal brief flashes in a two-alternative forced-choice task. **c**, Insensitivity to distance. Stimuli were enlarged according to a cortical magnification factor¹⁶ for easier judging. $n = 3$. **d**, Contrast polarity determines the direction of brightness change: the experiment shown in Fig. 1b was repeated with dark (less than 1 cd m^{-2}), equiluminant (grey flashes against a 5.91 cd m^{-2} green background), and light (11.82 cd m^{-2} light background). The luminance of the long flash was 5.91 cd m^{-2} ; the luminance of the brief flash ranged between $\pm 40\%$ of that of the long flash. The ordinate shows the percentage difference between the two conditions.

flashes at opposite edges of the computer screen (see Supplementary Information).

We also investigated the possibility that the offset-matched flash appears brighter (or perhaps more salient) because attention shifts to it. However, attention shifts are slow (~ 200 – 300 ms, ref. 8), whereas the brief flash appears and disappears within 56 ms. Nonetheless, we manipulated the contrast polarity of the background to test the attentional explanation. When the background was brighter than the flashes, the offset-matched brief flash appeared dimmer, not brighter, than the onset-matched flash (Fig. 2d; $t = -2.2$, $P = 0.026$). More importantly, when the background and flashes were made equiluminant, observers perceived no significant brightness illusion, arguing against a salience-enhancing attentional shift (Fig. 2d; $t = -0.58$, $P = 0.28$).

Spatial interactions in brightness perception are well known^{1,2} but provide an incomplete description; the TCE shows that brightness is influenced by temporal context as well. Our results further show a surprising juxtaposition of facts: first, that information encoded about the brightness of stimuli changes over time, such that the appearance of physically identical brief flashes compared to a persisting long flash varies as a function of stimulus onset asynchrony (Fig. 1c); and yet, second, the perceived brightness of a long flash remains constant over time (Fig. 2a). This indicates that brightness encoding might involve at least two neural populations: one with an adapting response that diminishes over time, and the other with a downstream response that assigns brightness labels to objects and does not adapt. We propose that the TCE arises from an interaction between these non-adapting and adapting encodings. In our model, activity in the non-adapting population remains constant—thereby encoding an unchanging label—even while its input from the adapting population diminishes (Supplementary Fig. S1; for an example of such hysteresis, see ref. 9). The hypothesis that some neurons maintain a brightness label is consistent with the idea that the brain strives to monitor the external world undistracted by predictable changes in its own physiology.

The activity of some neurons in primary visual cortex is correlated with brightness; that is, the firing rate varies as a function of surrounding luminance, even as the luminance in the receptive field remains unchanged^{10–14}. The activity of these brightness-responsive neurons adapts with time: the firing rates peak after 100 ms and then drop off precipitously¹⁴. This decrease in firing is consistent with an adapting population, although in some circumstances the later part of the spike train may correspond to the response of a non-adapting population¹⁴. This raises the possibility that the two encodings could be multiplexed into the same population of cells in V1. The TCE separates the physical measure of luminance and the perceptual quality of brightness (which typically co-vary) and offers a new approach for examining the neural coding of brightness. □

Received 30 September 2003; accepted 4 March 2004; doi:10.1038/nature02467.
Published online 14 April 2004.

- Gilchrist, A. *Lightness, Brightness and Transparency* (Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1994).
- Eagleman, D. M. Visual illusions and neurobiology. *Nature Rev. Neurosci.* **2**, 920–926 (2001).
- Koffka, K. *Principles of Gestalt Psychology* (Harcourt, Brace, and World, New York, 1935).
- Knill, D. C. & Kersten, D. Apparent surface curvature affects lightness perception. *Nature* **351**, 228–230 (1991).
- Adelson, E. H. Perceptual organization and the judgment of brightness. *Science* **262**, 2042–2044 (1993).
- Gilchrist, A. L. Lightness contrast and failures of constancy: a common explanation. *Percept. Psychophys.* **43**, 415–424 (1988).
- Broca, A. & Sulzer, D. La sensation lumineuse en fonction du temps. *J. Physiol. Pathol. Gén.* **4**, 632–640 (1902).
- Weichselgartner, E. & Sperling, G. Dynamics of automatic and controlled visual attention. *Science* **238**, 778–780 (1987).
- Hahnloser, R. H., Sarpeshkar, R., Mahowald, M. A., Douglas, R. J. & Seung, H. S. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature* **405**, 947–951 (2000).
- Rossi, A. F. & Paradiso, M. A. Neural correlates of perceived brightness in the retina, lateral geniculate nucleus, and striate cortex. *J. Neurosci.* **19**, 6145–6156 (1999).
- Rossi, A. F., Rittenhouse, C. D. & Paradiso, M. A. The representation of brightness in primary visual cortex. *Science* **273**, 1104–1107 (1996).

- MacEvoy, S. P., Kim, W. & Paradiso, M. A. Integration of surface information in primary visual cortex. *Nature Neurosci.* **1**, 616–620 (1998).
- Hung, C. P., Ramsden, B. M., Chen, L. M. & Roe, A. W. Building surfaces from borders in Areas 17 and 18 of the cat. *Vision Res.* **41**, 1389–1407 (2001).
- Kinoshita, M. & Komatsu, H. Neural representation of the luminance and brightness of a uniform surface in the macaque primary visual cortex. *J. Neurophysiol.* **86**, 2559–2570 (2001).
- Bowen, R. W., Pokorny, J. & Smith, V. C. Sawtooth contrast sensitivity: decrements have the edge. *Vision Res.* **29**, 1501–1509 (1989).
- Cowey, A. & Rolls, E. T. Human cortical magnification factor and its relation to visual acuity. *Exp. Brain Res.* **21**, 447–454 (1974).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank A. Holcombe, S. Anstis, X. Huang and M. Fallah for feedback. This research was supported by the Howard Hughes Medical Institute (D.M.E. and T.J.S.) and a grant from the Chapman Foundation and NSF IGERT (J.E.J.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.M.E. (david.eagleman@uth.tmc.edu) or J.E.J. (jacobson@salk.edu).

Integration of quanta in cerebellar granule cells during sensory processing

Paul Chadderton, Troy W. Margrie & Michael Häusser

Wolfson Institute for Biomedical Research and Department of Physiology, University College London, Gower Street, London WC1E 6BT, UK

To understand the computations performed by the input layers of cortical structures, it is essential to determine the relationship between sensory-evoked synaptic input and the resulting pattern of output spikes. In the cerebellum, granule cells constitute the input layer, translating mossy fibre signals into parallel fibre input to Purkinje cells¹. Until now, their small size and dense packing^{1,2} have precluded recordings from individual granule cells *in vivo*. Here we use whole-cell patch-clamp recordings to show the relationship between mossy fibre synaptic currents evoked by somatosensory stimulation and the resulting granule cell output patterns. Granule cells exhibited a low ongoing firing rate, due in part to dampening of excitability by a tonic inhibitory conductance mediated by GABA_A (γ -aminobutyric acid type A) receptors. Sensory stimulation produced bursts of mossy fibre excitatory postsynaptic currents (EPSCs) that summate to trigger bursts of spikes. Notably, these spike bursts were evoked by only a few quantal EPSCs, and yet spontaneous mossy fibre inputs triggered spikes only when inhibition was reduced. Our results reveal that the input layer of the cerebellum balances exquisite sensitivity with a high signal-to-noise ratio. Granule cell bursts are optimally suited to trigger glutamate receptor activation^{3–5} and plasticity^{6–8} at parallel fibre synapses, providing a link between input representation and memory storage in the cerebellum.

Cerebellar granule cells are the smallest and most numerous neurons in the mammalian brain and have a simple morphology with an average of only four short dendrites^{1,2}, each dendrite receiving a single excitatory mossy fibre input. Each granule cell also receives inhibitory input from one or two Golgi cells^{1,2}, the major granular layer interneuron (Fig. 1a). The combination of such a small number of inputs to an electrotonically compact structure^{9,10} provides an ideal model in which to study synaptic

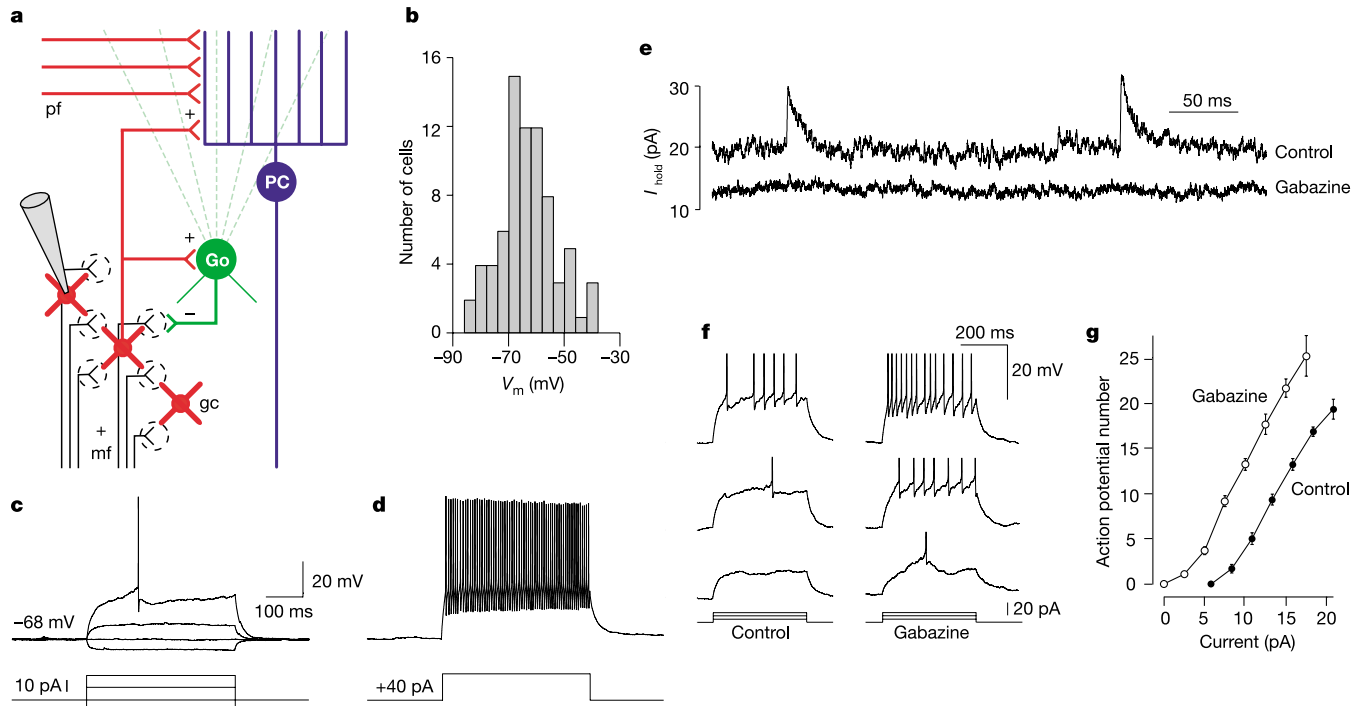


Figure 1 Intrinsic membrane properties of granule cells *in vivo*. **a**, Schematic diagram showing the basic connectivity in the cerebellar granule cell layer (mf, mossy fibre; gc, granule cell; pf, parallel fibre; PC, Purkinje cell; Go, Golgi cell). **b**, Histogram of granule cell resting membrane potentials (V_m ; ketamine/xylazine or urethane anaesthesia). **c**, **d**, Responses to current injection at resting potential show inward rectification and non-

accommodating spikes at up to 250 Hz (urethane anaesthesia). **e**, Voltage-clamp recording at 0 mV. Gabazine (0.5 mM) abolished outward currents and reduced the tonic holding current. **f**, Responses to current injection before and after gabazine application (action potentials truncated). **g**, Increased excitability in gabazine is reflected in a leftward shift of the f - I curve.

integration *in vivo*. To investigate input-output transformations during sensory activation, we made whole-cell patch-clamp recordings from granule cells in crus I and IIa of the cerebellar cortex, regions that produce strong multiunit and field responses during tactile stimulation of the vibrissae and the lip area in anaesthetized rats^{11–13}. Granule cells were identified by their depth ($>400\ \mu\text{m}$ from the pial surface) and their characteristic electrophysiological properties (high input resistance ($1.1 \pm 0.1\ \text{G}\Omega$ at rest, $n = 62$ cells), small capacitance ($\leq 5\ \text{pF}$) and fast membrane time constant ($\tau = 6.9 \pm 0.7\ \text{ms}$, $n = 49$; refs 9, 14, 15)). Granule cells exhibited a resting membrane potential of $-64 \pm 1\ \text{mV}$ ($24 \pm 1\ \text{mV}$ hyperpolarized to threshold, $n = 75$; Fig. 1b) and strong inward rectification (Fig. 1c), with input resistance nearly doubling when depolarizing from -75 to $-50\ \text{mV}$ (ratio 1.9 ± 0.1 , $n = 17$). Firing frequencies of up to 250 Hz were observed with little or no accommodation (accommodation ratio = 0.9 ± 0.1 , $n = 11$) in response to strong depolarizing current pulses (Fig. 1d). These properties are consistent with those described for granule cells *in vitro*^{14–17}.

Granule cells exhibited a low firing rate *in vivo*, with the mean spontaneous firing rate in the absence of holding current being $0.5 \pm 0.2\ \text{Hz}$ ($n = 46$). This low firing rate may reflect low excitatory drive or strong inhibition. To investigate synaptic input to granule cells *in vivo*, voltage-clamp recordings were made at -70 or $0\ \text{mV}$ (the reversal potentials for GABA_A and AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) receptor-mediated currents, respectively). At $-70\ \text{mV}$, spontaneous EPSCs sensitive to the AMPA receptor antagonist NBQX were recorded at a frequency of $4 \pm 1\ \text{Hz}$ (amplitude $-8.7 \pm 1.1\ \text{pA}$, $n = 10$). At $0\ \text{mV}$, spontaneous inhibitory postsynaptic currents (IPSCs) sensitive to the selective GABA_A receptor blocker gabazine (SR95531) were observed at a frequency of $5 \pm 1\ \text{Hz}$ (amplitude $-9.6 \pm 2.7\ \text{pA}$, $n = 10$). In addition to blocking IPSCs (Fig. 1e), gabazine also

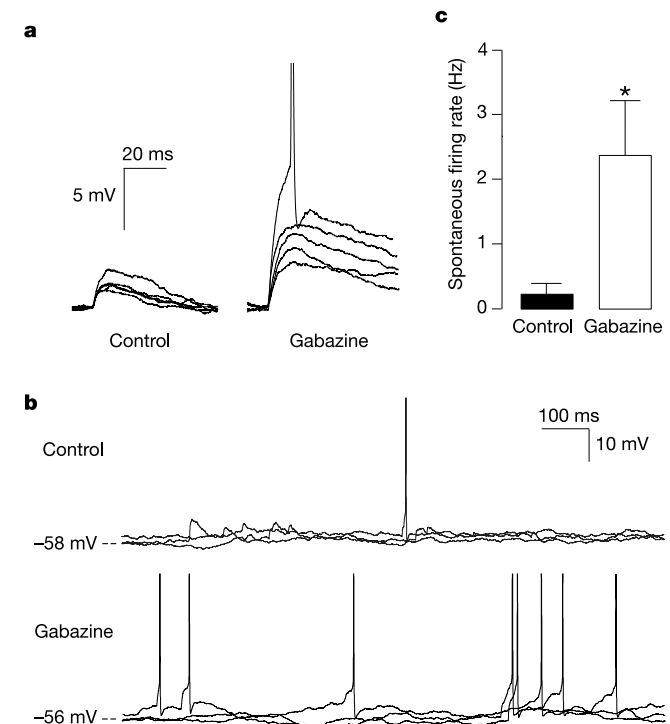


Figure 2 Low spontaneous firing rates are enforced by tonic inhibition *in vivo*.

a, Representative spontaneous EPSPs before and after GABA_A receptor blockade (action potential is truncated). EPSP amplitudes were increased in the presence of gabazine ($1.9 \pm 0.4\ \text{mV}$ to $3.7 \pm 0.5\ \text{mV}$, $P < 0.02$, $n = 4$). **b**, Current-clamp recording (three overlaid traces) from a granule cell at rest (top) and after GABA_A receptor blockade (bottom). **c**, Spontaneous firing rates increased in the presence of gabazine (asterisk, $P < 0.05$, $n = 5$).

produced a significant reduction in the baseline holding current ($P < 0.001$, $n = 11$), an increase in input resistance measured at -60 mV (1.3 ± 0.5 to 2.5 ± 0.4 G Ω , $P < 0.01$, $n = 6$) and a depolarization of resting potential (3.8 ± 1.4 mV, $P < 0.05$, $n = 7$). This indicates that a tonic GABA_A receptor-mediated conductance^{15,16,18,19}, due to persistent activation of extrasynaptic high-affinity GABA receptors^{15,20}, is present *in vivo*. This conductance had a powerful impact on the input-output function of granule cells *in vivo*, with gabazine producing a significant shift in the spike frequency-current (f - I) relationship ($P < 0.01$, $n = 3$; Fig. 1f, g) and increasing spontaneous mossy fibre excitatory postsynaptic potential (EPSP) amplitudes ($P < 0.02$, $n = 4$; Fig. 2a). As a consequence, during blocking of tonic inhibition, single EPSPs were on occasion capable of triggering action potentials (Fig. 2a; $n = 4$), with spontaneous firing rate increasing by an order of magnitude (Fig. 2b, c). Thus, spontaneous mossy fibre input is normally insufficient to trigger spike output owing in part to dampening of excitability by tonic inhibition. However, when tonic inhibition is reduced, unitary inputs are capable of driving granule cell output.

The input-output relationship of granule cells in response to physiologically relevant stimuli is unknown. We used air-puff stimulation of the vibrissae (or other facial regions; for example, upper lip) to evoke mossy fibre input to granule cells^{11–13}. Sensory stimulation evoked excitatory responses in 14 granule cells (only

one of which demonstrated both evoked EPSCs and IPSCs). The latency of EPSCs from the onset of the stimulus was 40 ± 6 ms (range = 9–75 ms, $n = 10$), consistent with a predominantly corticopontine origin of evoked mossy fibre activity¹³. Evoked EPSCs usually occurred in high-frequency bursts (defined as synaptic events with intercurrent intervals < 50 ms; average 3.2 ± 0.5 EPSCs at 75 ± 19 Hz, $n = 10$; see Fig. 3a, c, e). EPSCs within the burst were highly irregular (coefficient of variation of evoked EPSC intervals = 0.82 ± 0.07 , $n = 5$). Sensory stimulation produced action potentials in seven out of ten whole-cell recordings where evoked responses were observed, with the most common discharge pattern consisting of a burst of several action potentials (3.3 ± 0.2 action potentials, $n = 7$, including two cell-attached recordings; Fig. 3b, d, f). The mean frequency of action potentials within evoked bursts was 77 ± 12 Hz, with maximal instantaneous frequencies being as high as 250 Hz (mean = 163 ± 19 Hz, $n = 7$). Changing stimulus duration did not significantly affect the pattern of bursting responses (action potential number, $F_{(3,136)} = 2.3$, $P > 0.05$; action potential frequency, $F_{(3,97)} = 0.6$, $P > 0.05$; one-way analysis of variance, $n = 3$ cells for 30–300 ms duration).

To investigate the origin of individual evoked currents, we compared the amplitudes of spontaneous EPSCs, which presumably reflect release from single mossy fibre boutons, with those evoked by sensory stimulation. The similarity in amplitude distributions of these two populations of EPSCs (Fig. 4a, b; mean -8.9 ± 1.0 pA

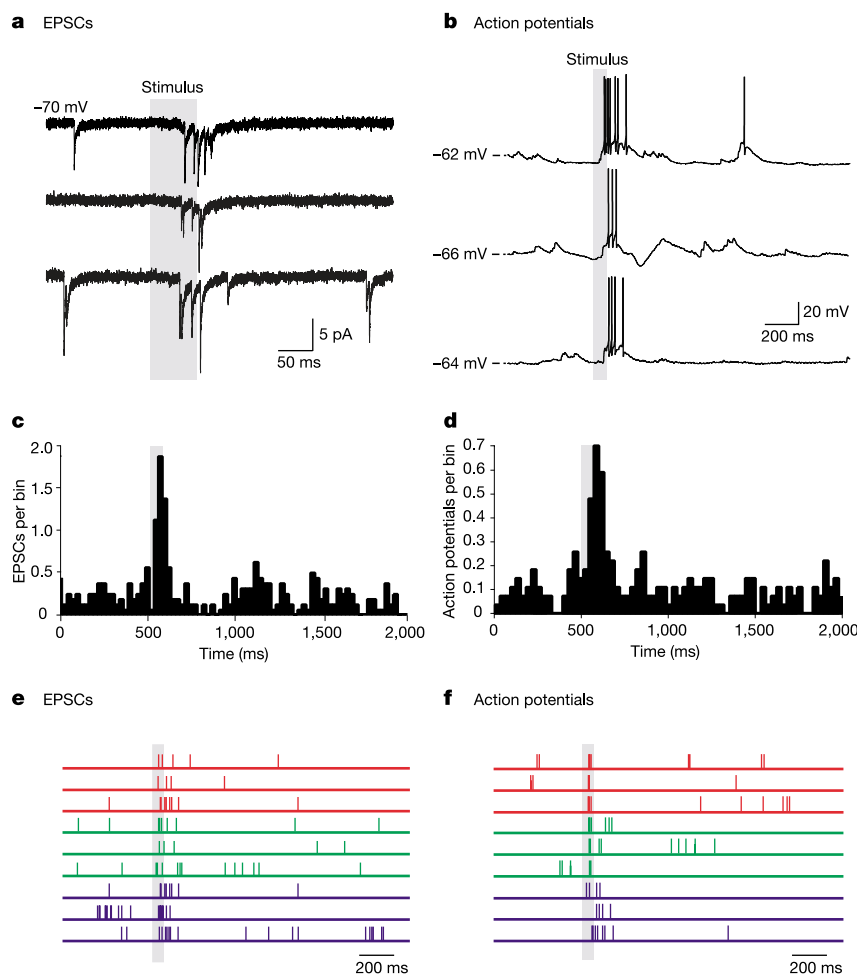


Figure 3 Evoked sensory responses in granule cells *in vivo*. **a**, EPSCs evoked by whisker stimulation recorded in voltage clamp at -70 mV (three consecutive trials). **b**, Action potentials evoked in response to whisker stimulation (three consecutive trials). **c**, **d**, Representative peristimulus time histogram of sensory-evoked EPSCs (**c**, 16 trials)

and action potentials (**d**, 16 trials) for individual granule cells. **e**, Raster plot of evoked EPSCs (three trials from three different granule cells, where each colour corresponds to a different cell). **f**, Raster plot of evoked action potentials (three trials from three different granule cells, with colours corresponding to the same cells as in **e**).

and -7.9 ± 0.6 pA at -70 mV for spontaneous and evoked EPSCs, respectively; $P = 0.28$, Kolmogorov–Smirnov test) indicates that individual evoked EPSCs originate from a single mossy fibre, with evoked EPSC bursts resulting from either a burst in a single presynaptic mossy fibre^{21,22} or asynchronous activation of multiple mossy fibres. To determine the contribution of single quanta, we isolated miniature EPSCs by applying tetrodotoxin (TTX). Remarkably, this did not significantly change the amplitude distribution of EPSCs compared to control (Fig. 4c, d; mean amplitudes -8.6 ± 0.4 pA and -8.1 ± 0.3 pA for spontaneous and miniature EPSCs, respectively; $n = 4$, $P = 0.77$, Kolmogorov–Smirnov test). The similarity of miniature spontaneous and evoked EPSCs indicates that release probability at mossy fibre synapses is likely to be low, and that sensory stimulation evokes bursts of EPSCs that each are composed of only one or two quanta.

To explore the link between the bursting pattern of mossy fibre EPSC inputs and bursting output, we compared spiking responses to different temporal patterns of EPSC waveforms injected into the granule cell *in vivo*. A burst of three EPSCs at 80 Hz produced a burst of action potentials that could not be evoked by a single EPSC with the same charge integral (2.9 ± 0.2 versus 1.5 ± 0.3 action potentials, respectively, $P < 0.001$, unpaired *t*-test, $n = 4$; Supplementary Fig. 1). This provides direct evidence that the tem-

porally distributed pattern of input promotes burst firing in granule cells. Burst firing will be enhanced further by the nonlinear properties of synaptic integration exhibited by granule cells, which include NMDA receptor and persistent sodium channel activation^{14,17}. Nonlinearities also contribute to integration *in vivo*, with depolarization from -60 mV to -50 mV producing a 1.6-fold increase in the integral of spontaneous EPSPs ($P < 0.05$, $n = 8$; Supplementary Fig. 2).

In a few granule cells, we were able to record both evoked EPSCs and action potentials in the same cells, allowing us to directly relate mossy fibre input to action potential output (Fig. 4e; $n = 5$). These experiments demonstrate that the number of output action potentials is intimately linked to the number of input EPSCs. On average, the EPSC/action potential relationship falls below the unity line, indicating that EPSP summation is required to generate spiking. This could be observed directly in some experiments, where action potentials usually occurred only after summation of two or more EPSPs (Fig. 4f).

We have taken advantage of the highly favourable anatomical and physiological properties of the cerebellar granule cell to investigate directly single-cell input–output computations in a cortical input layer *in vivo*. Our findings provide the first direct evidence in support of theoretical predictions that the spontaneous firing rate of granule cells *in vivo* is low, which will increase the number of distinct patterns of input that can be represented by the cerebellar cortex^{23,24}. Low firing rates are ensured by a hyperpolarized membrane potential and a tonic GABAergic conductance. We show that this conductance regulates the transformation of mossy fibre input into granule cell output, such that when tonic inhibition is reduced (as may occur after nitric oxide release²⁵ or changes in steroid levels²⁰), information arriving via a single mossy fibre action potential can be transferred through the granule cell layer. This is a critical requirement if the network is to remain sensitive to low levels of afferent activity²³.

Our finding that bursts of action potentials in granule cells are triggered by clustered sensory-evoked mossy fibre EPSCs complements and enhances existing theories of cerebellar function^{23,24} that propose that the granule cell layer is a low-noise sparse coding system. This burst-to-burst sequence ensures that granule cells reliably relay sensory-evoked mossy fibre signals, whereas events not associated with sensory stimulation are filtered out. Bursting thus further enhances the signal-to-noise ratio for transfer of sensory information in an input layer exhibiting low firing rates^{26–29}. We demonstrate that the bursting pattern of granule cell discharge is due to temporally distributed synaptic input. Notably, these clusters of EPSCs are composed of only a small number of quanta, indicating that granule cells are highly sensitive to individual quanta when they are temporally correlated by sensory input. This relay can therefore approach maximal sensitivity (that is, one quantum) depending on tonic inhibition and levels of mossy fibre input²³.

Bursts in granule cells have special relevance for the cerebellar network, as the downstream synapses made onto Purkinje cells show frequency-dependent facilitation of transmitter release³ and glutamate receptor activation⁴, suggesting that granule cell bursting will be nonlinearly amplified at the next synaptic relay²⁷. This will enhance postsynaptic activation of Purkinje cells through both ascending branch³⁰ and parallel fibre synapses, with the duration of the burst offering a relatively broad window for postsynaptic coincidence detection. Because induction of synaptic plasticity at these synapses is also highly frequency-dependent^{6–8}, granule cell bursting may provide a substrate for the storage of sensory representations. □

Methods

Patch-clamp recordings were made from granule cells in the cerebellar cortex of freely breathing 18–27-day-old Sprague–Dawley rats anaesthetized with urethane (1.2 g kg⁻¹) or with a ketamine (50 mg kg⁻¹)/xylazine (5 mg kg⁻¹) mixture as described²⁹. Current-

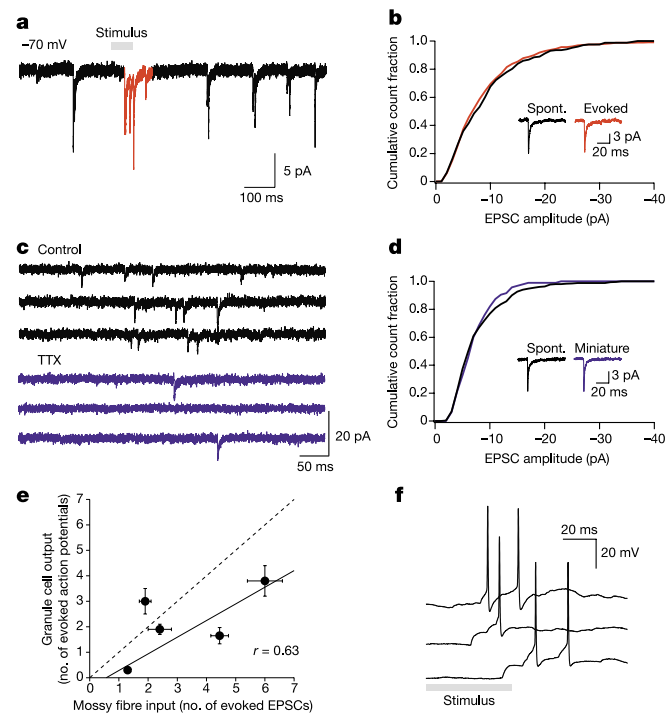


Figure 4 The relationship between mossy fibre input and granule cell output. **a**, Representative sensory-evoked EPSCs (red). **b**, Normalized cumulative amplitude histogram for spontaneous and evoked EPSCs ($n = 4$ cells; 1 pA bins). Inset: average EPSCs from a single cell. **c**, Spontaneous EPSCs and miniature EPSCs after TTX application. **d**, Normalized cumulative amplitude histogram for spontaneous and miniature EPSCs ($n = 5$, 1 pA bins). Inset: average EPSCs from a single cell. **e**, The relationship between the number of evoked mossy fibre EPSCs and granule cell action potentials at resting potential (-55 ± 5 mV). Each point represents a single granule cell. The dashed line indicates when the EPSC/action potential ratio equals one. Cells below this line (4 out of 5) require EPSP summation in order to generate output. The solid line is a linear fit to data points weighted by s.e.m. **f**, Representative traces showing summation of evoked EPSPs and resulting action potentials. In 86% of trials, at least two sensory-evoked EPSPs were required to summate to produce the first action potential in the response.

clamp and voltage-clamp recordings were made using a Multiclamp 700A amplifier (Axon Instruments). Patch pipettes (4–8 M Ω) were filled with a K-methanesulphonate-based internal solution (7 mM Cl[−]). Sensory responses were evoked by an air puff (30–300 ms, 60 p.s.i.) timed by a Picospritzer (General Valve) and delivered to the ipsilateral perioral surface. Data are represented as mean $\pm$ s.e.m. See Supplementary Methods for further details.

Received 12 January; accepted 26 February 2004; doi:10.1038/nature02442.

1. Eccles, J. C., Ito, M. & Szentagothai, J. *The Cerebellum as a Neuronal Machine* (Springer, Berlin, 1967).
2. Jakab, R. L. & Hamori, J. Quantitative morphology and synaptology of cerebellar glomeruli in the rat. *Anat. Embryol. (Berl.)* **179**, 81–88 (1988).
3. Konnerth, A., Llano, I. & Armstrong, C. M. Synaptic currents in cerebellar Purkinje cells. *Proc. Natl Acad. Sci. USA* **87**, 2662–2665 (1990).
4. Takechi, H., Eilers, J. & Konnerth, A. A new class of synaptic response involving calcium release in dendritic spines. *Nature* **396**, 757–760 (1998).
5. Perkel, D. J., Hestrin, S., Sah, P. & Nicoll, R. A. Excitatory synaptic currents in Purkinje cells. *Proc. R. Soc. Lond. B* **241**, 116–121 (1990).
6. Casado, M., Isope, P. & Ascher, P. Involvement of presynaptic N-methyl-D-aspartate receptors in cerebellar long-term depression. *Neuron* **33**, 123–130 (2002).
7. Wang, S. S.-H., Denk, W. & Häusser, M. Coincidence detection in single dendritic spines mediated by calcium release. *Nature Neurosci.* **3**, 1266–1273 (2000).
8. Brown, S. P., Brenowitz, S. D. & Regehr, W. G. Brief presynaptic bursts evoke synapse-specific retrograde inhibition mediated by endogenous cannabinoids. *Nature Neurosci.* **6**, 1048–1057 (2003).
9. Silver, R. A., Traynelis, S. F. & Cull-Candy, S. G. Rapid-time-course miniature and evoked excitatory currents at cerebellar synapses *in situ*. *Nature* **355**, 163–166 (1992).
10. Gabbiani, F., Midtgard, J. & Knöpfel, T. Synaptic integration in a model of cerebellar granule cells. *J. Neurophysiol.* **72**, 999–1009 (1994).
11. Shambes, G. M., Gibson, J. M. & Welker, W. Fractured somatotopy in granule cell tactile areas of rat cerebellar hemispheres revealed by micromapping. *Brain Behav. Evol.* **15**, 94–140 (1978).
12. Bower, J. M. & Woolston, D. C. Congruence of spatial organization of tactile projections to granule cell and Purkinje cell layers of cerebellar hemispheres of the albino rat: vertical organization of cerebellar cortex. *J. Neurophysiol.* **49**, 745–766 (1983).
13. Morissette, J. & Bower, J. M. Contribution of somatosensory cortex to responses in the rat cerebellar granule cell layer following peripheral tactile stimulation. *Exp. Brain Res.* **109**, 240–250 (1996).
14. D'Angelo, E., De Filippi, G., Rossi, P. & Taglietti, V. Synaptic excitation of individual rat cerebellar granule cells *in situ*: evidence for the role of NMDA receptors. *J. Physiol. (Lond.)* **484**, 397–413 (1995).
15. Brickley, S. G., Revilla, V., Cull-Candy, S. G., Wisden, W. & Farrant, M. Adaptive regulation of neuronal excitability by a voltage-independent potassium conductance. *Nature* **409**, 88–92 (2001).
16. Brickley, S. G., Cull-Candy, S. G. & Farrant, M. Development of a tonic form of synaptic inhibition in rat cerebellar granule cells resulting from persistent activation of GABA_A receptors. *J. Physiol. (Lond.)* **497**, 753–759 (1996).
17. D'Angelo, E., De Filippi, G., Rossi, P. & Taglietti, V. Ionic mechanism of electroresponsiveness in cerebellar granule cells implicates the action of a persistent sodium current. *J. Neurophysiol.* **80**, 493–503 (1998).
18. Wall, M. J. & Usowicz, M. M. Development of action potential-dependent and independent spontaneous GABA_A receptor-mediated currents in granule cells of postnatal rat cerebellum. *Eur. J. Neurosci.* **9**, 533–548 (1997).
19. Hamann, M., Rossi, D. J. & Attwell, D. Tonic and spillover inhibition of granule cells control information flow through cerebellar cortex. *Neuron* **33**, 625–633 (2002).
20. Stell, B. M., Brickley, S. G., Tang, C. Y., Farrant, M. & Mody, I. Neuroactive steroids reduce neuronal excitability by selectively enhancing tonic inhibition mediated by δ subunit-containing GABA_A receptors. *Proc. Natl Acad. Sci. USA* **100**, 14439–14444 (2003).
21. Eccles, J. C., Faber, D. S., Murphy, J. T., Sabah, N. H. & Taborikova, H. Afferent volleys in limb nerves influencing impulse discharges in cerebellar cortex. I. In mossy fibers and granule cells. *Exp. Brain Res.* **13**, 15–35 (1971).
22. Garwicz, M., Jörntell, H. & Ekerot, C. F. Cutaneous receptive fields and topography of mossy fibres and climbing fibres projecting to cat cerebellar C3 zone. *J. Physiol. (Lond.)* **512**, 277–293 (1998).
23. Marr, D. A theory of cerebellar cortex. *J. Physiol. (Lond.)* **202**, 437–470 (1969).
24. Albus, J. S. A theory of cerebellar function. *Math. Biosci.* **10**, 25–61 (1971).
25. Wall, M. J. Endogenous nitric oxide modulates GABAergic transmission to granule cells in adult rat cerebellum. *Eur. J. Neurosci.* **18**, 869–878 (2003).
26. Krahe, R. & Gabbiani, F. Burst firing in sensory systems. *Nature Rev. Neurosci.* **5**, 13–23 (2004).
27. Lisman, J. E. Bursts as a unit of neural information: making unreliable synapses reliable. *Trends Neurosci.* **20**, 38–43 (1997).
28. Hahnloser, R. H., Kozhevnikov, A. A. & Fee, M. S. An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* **419**, 65–70 (2002).
29. Margrie, T. W., Brecht, M. & Sakmann, B. *In vivo*, low-resistance, whole-cell recordings from neurons in the anaesthetized and awake mammalian brain. *Pflügers Arch.* **444**, 491–498 (2002).
30. Llinás, R. in *The Cerebellum: New Vistas* (eds Palay, S. L. & Chan-Palay, V.) 189–194 (Springer, New York, 1982).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank B. Clark, J. Davie, M. Farrant and A. Roth for comments. This work was supported by grants from the European Union (M.H.), Wellcome Trust (M.H. and T.W.M.), Gatsby Foundation (M.H.), NHMRC (T.W.M.) and by a UCL Graduate School Research Scholarship (P.C.). T.W.M. acknowledges the MPI für medizinische Forschung, Heidelberg, for support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.H. (m.hauser@ucl.ac.uk).

Birth of parthenogenetic mice that can develop to adulthood

Tomohiro Kono^{1,3}, Yayoi Obata^{1,3}, Quiong Wu^{1,3}, Katsutoshi Niwa^{1,3}, Yukiko Ono¹, Yuji Yamamoto^{2,3}, Eun Sung Park⁴, Jeong-Sun Seo^{4,5} & Hidehiko Ogawa^{1,3}

¹Department of BioScience, and ²Department of Applied Science, Tokyo University of Agriculture, Setagaya-ku, Tokyo 156-8502, Japan

³Bio-oriented Technology Research Advancement Institution (BRAIN), Minato-ku, Tokyo 105-0001, Japan

⁴MacroGen Inc, Chongno-Ku, Seoul 110-061, Korea

⁵Department of Biochemistry, Seoul National University College of Medicine, Chongno-Ku, Seoul 110-799, Korea

Only mammals have relinquished parthenogenesis, a means of producing descendants solely from maternal germ cells. Mouse parthenogenetic embryos die by day 10 of gestation^{1–4}. Biparental reproduction is necessary because of parent-specific epigenetic modification of the genome during gametogenesis^{5–8}. This leads to unequal expression of imprinted genes from the maternal and paternal alleles⁹. However, there is no direct evidence that genomic imprinting is the only barrier to parthenogenetic development. Here we show the development of a viable parthenogenetic mouse individual from a reconstructed oocyte containing two haploid sets of maternal genome, derived from non-growing and fully grown oocytes. This development was made possible by the appropriate expression of the *Igf2* and *H19* genes with other imprinted genes, using mutant mice with a 13-kilobase deletion in the *H19* gene¹⁰ as non-growing oocytes donors. This full-term development is associated with a marked reduction in aberrantly expressed genes. The parthenote developed to adulthood with the ability to reproduce offspring. These results suggest that paternal imprinting prevents parthenogenesis, ensuring that the paternal contribution is obligatory for the descendant.

Maternal-specific *de novo* methylation of imprinted loci occurs during oocyte growth^{7,8,11,12}; such non-growing oocytes from newborn mice are considered to be naive with respect to most of the maternal imprinting process. Parthenogenetic mouse embryos (ng^{wt}/fg^{wt}) that contain genomes from non-growing (ng^{wt}) and fully grown (fg^{wt}) oocytes can develop into 13.5-day-old fetuses¹², by the appropriate expression of many imprinted genes¹³. However, the imprinted expression of the *H19* and *Igf2* genes has not been altered in the ng allele. The expression of the *Igf2* and *H19* genes are coordinately regulated by *cis*-acting elements, depending on the methylation status of the differentially methylated region (DMR) of the *H19* gene^{14–16} and on endoderm-specific enhancers¹⁷ (Fig. 1i). To block *H19* gene expression from ng oocyte alleles, mutant mice harbouring a 3-kilobase (kb) deletion in the *H19* transcription unit¹⁸ were used to construct parthenogenetic embryos. The ng^{H19Δ3}/fg^{wt} parthenotes developed as live fetuses for 17.5 days of gestation, but showed no sign of further development¹⁹. The *Igf2*

Table 1 Development of reconstructed parthenogenetic embryos

Developmental progress	Number
Number of reconstructed eggs	457
Number of embryos developed to blastocysts	417 (91.2% of reconstructed eggs)
Number of embryos transferred	371 (89.0% of blastocysts)
Number pregnant/recipients	24/26
Number of implantation to recipients	246 (71.7% of embryos transferred to pregnant)
Number of pups	28 (8.2% of embryos transferred to pregnant)
Dead	18 (5.2% of embryos transferred to pregnant)
Live	8 (2.3% of embryos transferred to pregnant)
Survived	2 (0.6% of embryos transferred to pregnant)

gene, which encodes a growth-promoting factor (IGF-II), is well known as a major regulator of fetal growth^{10,18,20}. These experiments suggest that the expression of *Igf2* along with monoallelic expression of the *H19* gene leads to further development of parthenotes.

First we investigated the development of the $ng^{H19\Delta13}/fg^{wt}$ parthenogenetic embryos, which were expressing the *Igf2* gene from the *ng* allele. A total of 598 $ng^{H19\Delta13}/fg^{wt}$ oocytes were constructed by serial nuclear transfer¹² using *ng* oocytes from the *H19* ^{$\Delta13$} newborn mice¹⁰. After *in vitro* maturation and artificial activation, 78% of the oocytes resulted in diploid one-cell parthenogenetic embryos that formed two second polar bodies and pronuclei. Of the 457 eggs cultured *in vitro*, 91.2% developed to the morula/blastocyst stage (Table 1). In all, 371 morula/blastocysts derived from $ng^{H19\Delta13}/fg^{wt}$ oocytes were transferred to 26 recipient females, and of these 24 became pregnant. Beyond our expectations, a total of 10 live and 18 dead pups were recovered by autopsy at 19.5 days of gestation (Fig. 1a–d). Of these, two individual pups were successfully restored, showing the apparently normal neonate morphology (Fig. 1a, b). To confirm that this unique phenomenon is seen only in $ng^{H19\Delta13}/fg^{wt}$ parthenotes, we reassessed the development of $ng^{H19\Delta13}/fg^{wt}$ parthenotes¹⁹. The data confirmed our previous results that $ng^{H19\Delta13}/fg^{wt}$ parthenotes do not develop beyond day 17.5 of gestation (data not shown).

The body weights of the two surviving parthenogenetic pups were 1,372 and 1,310 mg (Fig. 1a, b), similar to those of the control

B6D2F1 $\times$ B6^{H19 $\Delta13$} pups (mean $\pm$ s.e.m, $1,326 \pm 30$ mg, $n = 12$). Polymerase chain reaction (PCR) analysis using specific primers for the *H19* gene and the *neo* cassette sequence revealed that they originated from $ng^{H19\Delta13}/fg^{wt}$ embryos (Fig. 1h). One of the two survivors was nursed by a foster mother, and grew to adulthood. She was named 'Kaguya' (Fig. 1a, e) and showed normal reproductive performance: after mating she conceived and delivered normal pups (Fig. 1e). The other survivor was used for gene expression analysis on the day of recovery. The remaining live (Fig. 1c) and dead (Fig. 1d) pups at recovery showed developmental retardation, the live pups dying within 15 min. The mean body weight (Fig. 1f) was 744 ± 25 mg ($n = 8$) in live pups and 786 ± 20 mg ($n = 14$) in dead pups, and the developmental stage was estimated to be stage 26, as represented by day 18 fetuses²¹.

Histological analysis showed immaturity of the liver in the growth-retarded $ng^{H19\Delta13}/fg^{wt}$ parthenotes. The sinusoids were packed with blood precursors, such as megakaryocytes, and erythroid and myeloid precursors, but the hepatic vein was not well differentiated (data not shown). The placenta was afflicted with atrophy (Fig. 1g) and the placental weights were 77 and 103 mg for the two survivors. However, gross abnormal formations of the spongiotrophoblast and labyrinth layer¹⁹ were not observed, suggesting that the placenta was functionally capable of supporting embryo development.

To explore and compare features of the parthenotes, we carried out global gene expression analysis by oligonucleotide mouse 11K

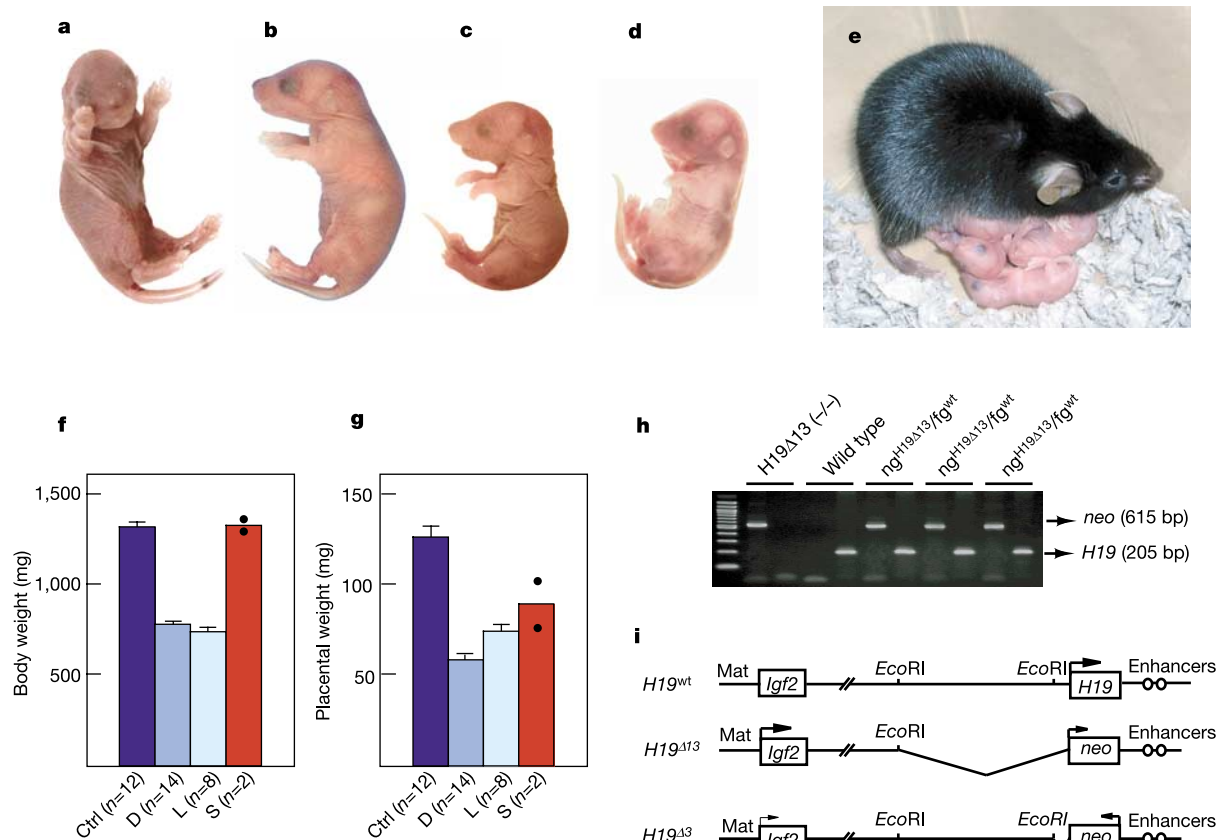


Figure 1 Parthenogenetic mice derived from $ng^{H19\Delta13}/fg^{wt}$ embryos. **a, b**, Parthenogenetic pups at birth with normal phenotypes. Of these the pup shown in **a** grew up a normal adult with reproductive competence (see **e**), and that in **b** was used for gene expression analysis. **c, d**, Growth-retarded parthenotes at day 19.5 of gestation, which died soon after the pup shown in **c** and just before that in **d**. **f, g**, Body (**f**) and placental (**g**) weights of the parthenotes at birth (Ctrl, controls $n = 12$; D, dead at birth $n = 14$; L, alive at birth $n = 8$; S, survived $n = 2$). Values are means $\pm$ s.e.m. Dots on

column S represent the individual weights of the survivors. **h**, Genotype analysis by PCR showing the parthenotes containing an *ng* allele in which the *H19* gene was deleted with the *neo* cassette and an *fg* allele of the wild type. **i**, Schematic representation of the *H19-Igf2* locus in *H19*^{wt}, *H19* ^{$\Delta13$} (ref. 10) and *H19* ^{$\Delta3$} (ref. 18) mutants. Only maternal alleles of each mutant are shown. The *H19* ^{$\Delta13$} mutant mice were used in the present study and the *H19* ^{$\Delta3$} mutants were used in our previous manuscript. The *EcoRI* fragment (-10 kb to -50 base pairs) contains the differentially methylated region.

microarray at day 12.5 of gestation. The phenotypes were similar to those of the control fetuses (Fig. 2a–c), although the fetal and placental weights of parthenotes were significantly lower than those of the controls (Fig. 2g, h). Interestingly, the umbilical cord was of normal thickness in the $ng^{H19\Delta13}/fg^{wt}$ parthenotes, but was very poor in the ng^{wt}/fg^{wt} parthenotes (Fig. 2d–f). The proportion of embryos developing to day 12.5 of gestation was not significantly different between $ng^{H19\Delta13}/fg^{wt}$ (35%, 16/46) and ng^{wt}/fg^{wt} (25%, 14/56) embryos. Clustering of the expression data and imaging were done using the Cluster and TreeView programs²², respectively. The gene expression profiles of the $ng^{H19\Delta13}/fg^{wt}$ parthenotes differed from that of the ng^{wt}/fg^{wt} parthenotes, over a wide range of genes (Supplementary Fig. S1). For 1,038 genes, statistical significance was shown at the nominal significance level ($P < 0.001$) of each univariate test. By ontology comparison, these genes can be classified into three categories: cell communication (15.1%), cell growth/maintenance (19.1%) and metabolism (21.9%) (Supplementary Fig. S2). These results suggest that the extended development in

the $ng^{H19\Delta13}/fg^{wt}$ parthenotes is the product of a wide-ranging alteration of gene expression.

The oligonucleotide microarray analysis also showed that in the $ng^{H19\Delta13}/fg^{wt}$ parthenotes, normalization occurred in all of the imprinted genes analysed (Supplementary Fig. S3). Out of 34 imprinted genes, only two genes, *Grb10* and *Nnat*, were respectively up- and downregulated at more than double and less than half the level in controls. However, 11 of the genes were downregulated and one gene, *Grb10*, was upregulated in the ng^{wt}/fg^{wt} parthenotes. Furthermore, the number of genes that were expressed differentially at a level greater than twofold ($\log_2$ (Cy5/Cy3) ratio $> +1$ and < -1) compared with controls for each fetus, ranged from only 11 to 42 (average 30) in the $ng^{H19\Delta13}/fg^{wt}$ parthenotes (Supplementary Fig. S4). In contrast, in the ng^{wt}/fg^{wt} parthenotes, a remarkably large number of genes, ranging from 431 to 1,324 (average 842), showed differential expression. Of these, *Dlk1* is the only gene that displayed changes in expression in all four $ng^{H19\Delta13}/fg^{wt}$ parthenotes, whereas a common set of 329 genes with varied expression was found in the ng^{wt}/fg^{wt} parthenotes. When $ng^{H19\Delta13}/fg^{wt}$ and ng^{wt}/fg^{wt} parthenotes were compared with each other, from 131 to 295 (average 208) genes displayed greater than twofold changes. Thus, the data clearly show that a marked normalization of expression of a wide-range of genes occurred in the $ng^{H19\Delta13}/fg^{wt}$ parthenotes, and the more physiological pattern of gene expression provided sufficient developmental competence for these parthenotes.

To address the question of which genes have a key role in the survival of the reconstructed parthenotes, we focused on two pairs of interrelated genes: *Igf2* and *H19* (refs 10, 14–18), and *Dlk1* and *Gtl2* (refs 23–27) (Fig. 3), which are imprinted during spermatogenesis. Quantitative gene expression analysis by real-time RT–PCR (PCR with reverse transcription) showed that both *Igf2* and *H19*

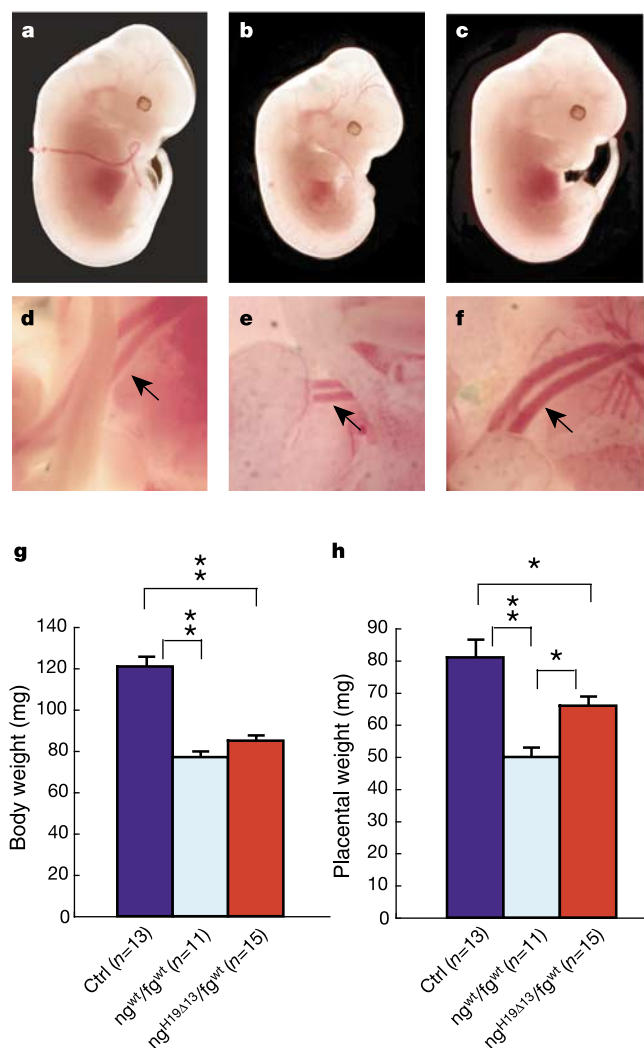


Figure 2 Parthenogenetic fetuses at day 12.5 of gestation. **a, d**, Fetuses and the umbilical cords derived from control fertilized embryos; **b, e**, ng^{wt}/fg^{wt} parthenogenetic embryos; **c, f**, $ng^{H19\Delta13}/fg^{wt}$ parthenogenetic embryos. Note that the umbilical cord (indicated by arrows) was well developed in the $ng^{H19\Delta13}/fg^{wt}$ fetus (**f**) but was very poor in the ng^{wt}/fg^{wt} parthenote. **g, h**, Graphical representations of body (**g**) and placental (**h**) weights. Values are means $\pm$ s.e.m. Significant differences at * $P < 0.05$ and ** $P < 0.01$.

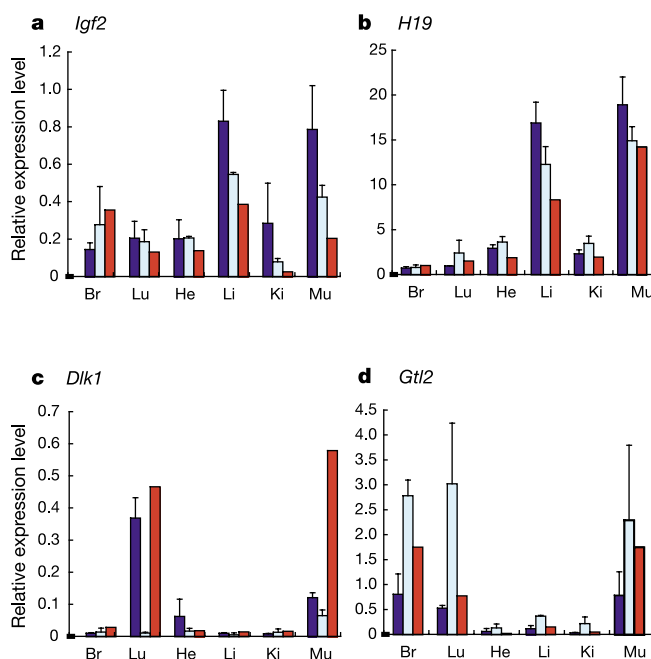


Figure 3 Graphical representations of the expression of the two sets of coordinate imprinted genes in parthenogenetic fetuses. The expressions of four imprinted genes, **a**, *Igf2*; **b**, *H19*; **c**, *Dlk1*; and **d**, *Gtl2* in six different tissues of 19.5-day $ng^{H19\Delta13}/fg^{wt}$ fetuses were analysed by quantitative real-time PCR. Br, brain; Lu, lung; He, heart; Li, liver; Ki, kidney; Mu, muscle. The values represent the levels of expression relative to that of the internal control gene (*Gapdh*). Control ($n = 3$; dark blue); growth-retarded $ng^{H19\Delta13}/fg^{wt}$ parthenogenetic pups ($n = 3$; light blue); and survived $ng^{H19\Delta13}/fg^{wt}$ parthenogenetic pup ($n = 1$; red). Values are means $\pm$ s.e.m.

genes were expressed at similar levels in the major organs of the ng^{H19Δ13}/fg^{wt} parthenote, as expected, because the *H19* transcription unit and its DMR had been deleted in the ng allele¹⁰. Interestingly, the *Gtl2* gene in the growth-retarded parthenotes was expressed at a level two- to fourfold higher than in the controls in all tissues analysed. Analysis using a DNA polymorphism present in the alleles of JF1 (*Mus musculus molossinus*) mice revealed that the *Gtl2* gene was biallelically expressed (data not shown). The *Dlk1* gene, which was expressed mainly in the lung, heart and muscle in the controls, was expressed in the lung and muscle in the ng^{H19Δ13}/fg^{wt} parthenotes. However, in the growth-retarded parthenotes, the expression in the lung was repressed. These expression patterns were confirmed by *in situ* hybridization at day 12.5 of gestation (data not shown). The decrease in *Gtl2* gene expression and activation of the *Dlk1* gene may be one of the causes of normal development of the survivor parthenotes. These genes are known to cause fetal lethality in the mutants²⁵. Moreover, the maternal disomy mice for distal chromosome 12, where *Dlk1* and *Gtl2* are located, die perinatally with growth retardation^{26,27}. Therefore, appropriate expression of *Dlk1* and *Gtl2* genes^{23–25} may lead to even greater improvements in the efficiency of parthenogenetic development in mice.

Our study shows that imprinting is a barrier to parthenogenetic development in mice. This is consistent with the 'parental conflict hypothesis' (ref. 28), which proposes opposite effects of the parental genomes, such as in growth regulation during development. By increasing the activity of the *Igf2* gene in parthenogenetic embryos together with monoallelic expression of the *H19* gene, we have shown for the first time that it is possible to obtain a viable adult mouse from two maternal genomes. The most fascinating riddle arising from these results is how appropriate expression of the *Igf2* and *H19* genes caused the modification of a wide range of genes and normal development in the ng^{H19Δ13}/fg^{wt} parthenotes. The better development might lead to altered signalling, which then affects the expression of many other genes; however, the underlying mechanism is still unclear. Nevertheless, this study emphasizes the fact that normal development in mice is subject to a rigorous 'conflict' and differences due to imprinting of maternal and paternal genomes^{29,30}. □

Methods

Oocyte manipulations

Fully-grown germinal vesicle (GV) oocytes were collected from the ovarian follicles of B6D2F1 (C57BL/6J × DBA) females 44–48 h after an equine chorionic gonadotrophin (eCG) injection. Ovulated MII oocytes were also collected from superovulated B6D2F1 mice 16 h after human chorionic gonadotrophin (hCG) injection. Mice harbouring a deletion of the *H19* transcription unit with DMR¹⁰ (total 13 kb), which were kindly donated by S. Tilghman, were used as ng-stage oocyte donors. The mutant mice were originally produced by injection of targeted CCE ES cells into C57BL/6J blastocysts, and the founder mutants were backcrossed with C57BL/6J females. Ng oocytes, which were at the diplotene stage of the first meiosis, were collected from the ovaries of 1-day-old newborn mice.

Nuclear transfer was carried out as outlined previously^{12,19}. Oocytes containing a haploid set of genomes from *H19*^{Δ13} null mutant females and another genome from ovulated MII oocytes of wild-type B6D2F1 mice were constructed by serial nuclear transfer. Fusion of the diplotene oocytes with enucleated GV oocytes was induced with an inactivated Sendai virus (HVJ, 2,700 haemagglutinating activity units per ml). After fusion, the reconstituted oocytes were cultured for 14 h in α-MEM medium (Gibco). The GV oocytes were manipulated in a medium containing 200 μM dbcAMP and 5% calf serum throughout the experiment and were released from this medium 1 h after fusion with a ng-stage oocyte. A set of MII chromosomes from the reconstituted oocytes was again transferred into an enucleated MII oocyte. Then the reconstructed oocytes were artificially activated with 10 mM SrCl₂ in Ca²⁺-free M16 medium for 2 h. These embryos were cultured in M16 medium for 3 days in an atmosphere of 5% CO₂, 5% O₂ and 90% N₂ at 37 °C. Blastocysts obtained from the constructed oocytes were transferred into the uterine horns of CD-1 female mice at 2.5 days of pseudopregnancy. The postimplantation development was assessed by an autopsy performed at 19.5 days of gestation.

Oligonucleotide microarray analysis

Briefly, RNA was prepared from whole embryos using the RNeasy Midi Kit according to the manufacturer's instructions (Qiagen). First-strand complementary DNA synthesis from 5 μg total RNA was done at 42 °C for 1 h, followed by second-strand cDNA synthesis

at 16 °C for 2 h. RNase I digestion was followed by proteinase K treatment. The resulting double-stranded cDNA was purified using the RNeasy Mini Kit according to the manufacturer's instructions (Qiagen). Fluorescence-labelled RNA was generated by carrying out an *in vitro* transcription reaction (40 mM Tris-HCl, pH 7.5, 7 mM MgCl₂, 10 mM NaCl, 2 mM spermidine, 5 mM dithiothreitol, 7.5 mM each of ATP, CTP and GTP, 5 mM UTP, 20 units RNase inhibitor and 1,000 units T7 RNA polymerase) at 37 °C for 4 h. This reaction was done in the presence of fluorescence-labelled nucleotides (Amersham Pharmacia Biotech) to generate Cy3- or Cy5-labelled RNA. The labelled RNA was subsequently purified (Qiagen RNeasy Mini Kit) and chemically fragmented at 94 °C for 15 min in fragmentation buffer (20 mM Tris-acetate, pH 8.1, 50 mM potassium acetate, 15 mM magnesium acetate). The fragmented, Cy3- or Cy5-labelled complementary RNA (20 μg) was lyophilized, resolubilized in hybridization buffer (50% formamide, 50 mM sodium phosphate, pH 8.0, 6 × SSC, 5 × Denhardt's Solution, 0.5% SDS), and hybridized to oligonucleotide mouse 11K microarrays (Macrogen) at 42 °C for 24 h. T7 RNA polymerase was purchased from Ambion; primers, RNase inhibitors and all other enzymes were purchased from Roche Molecular Biochemicals.

The Mouse 11K microarrays are spotted microarrays consisting of 50-mer oligonucleotide probes that represent 11,376 mouse genes, including 8,996 known and 2,380 unknown genes (<http://www.macrogen.com/>). Arrays were washed in three consecutive steps with 2 × SSC/0.1% SDS, 1 × SSC and 0.5 × SSC and scanned. Scans were performed on a Generation III scanner (Amersham Pharmacia Biotech), and the expression value for each gene was calculated using Image 5.0 software (BioDiscovery Inc.). Minor differences in microarray intensities were normalized using the SMA R package (Statistics for Microarray Analysis <http://www.stat.berkeley.edu/users/terry/zarray/Software/smcode.html>) developed by S. Dudoit, Y. H. Yang, M. Callow and T. Speed. Analyses were performed using Cluster developed by M. Eisen²⁷ and BRB ArrayTools developed by R. Simon and A. Peng (<http://linus.nci.nih.gov/BRB-ArrayTools.html>).

Quantitative gene expression analysis

Quantitative analysis of gene expression was performed by means of real-time PCR (LightCycler System, Roche Molecular Biochemicals) using a ready-to-use reaction mixture kit (LightCycler FirstStart DNA Master SYBR Green I, Roche Molecular Biochemicals). The primers used were as follows: *Igf2*, 5'-AGGGGAGCTTGTGACACG-3' and 5'-GGGTATCTGGGGAAGTCGTC-3'; *H19*, 5'-CATGTCTGGGCCTTTGAA-3' and 5'-TTGGCTCCAGGATGATGT-3'; *Dlk1*, 5'-ACTTGGCTGGACCTGGAGAA-3' and 5'-CTGTTGGTTGCGGCTACGAT-3'; *Gtl2*, 5'-AAGCACCATTAGCCACTAGG-3' and 5'-TTGCACATTTCCTCTGGGAC-3'; *Gapdh*, 5'-GTCTGGAGTCTACTGGTGTC-3' and 5'-GAGCCCTTCCACAATGCCAAA-3'.

Received 20 November 2003; accepted 10 February 2004; doi:10.1038/nature02402.

- Surani, M. A. H. & Barton, S. C. Development of gynogenetic eggs in the mouse: Implications for parthenogenetic embryos. *Science* **222**, 1034–1036 (1983).
- Surani, M. A. H., Barton, S. C. & Norris, M. L. Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* **308**, 548–550 (1984).
- Surani, M. A. *et al.* Genome imprinting and development in the mouse. *Development (suppl.)*, 89–98 (1990).
- McGrath, J. & Solter, D. Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* **37**, 179–183 (1984).
- Brannan, C. I. & Bartolomei, M. S. Mechanisms of genomic imprinting. *Curr. Opin. Genet. Dev.* **9**, 164–170 (1999).
- Tilghman, S. M. The sins of the fathers and mothers: Genomic imprinting in mammalian development. *Cell* **96**, 185–193 (1999).
- Mann, J. R. Imprinting in the germ line. *Stem Cells* **19**, 287–294 (2001).
- Li, E. Chromatin modification and epigenetic reprogramming in mammalian development. *Nature* **375**, 662–673 (2002).
- Obata, Y. & Kono, T. Maternal primary imprinting is established at a specific time for each gene throughout oocyte growth. *J. Biol. Chem.* **277**, 5285–5289 (2002).
- Leighton, P. A., Ingram, R. S., Eggenschwiler, J., Efstratiadis, A. & Tilghman, S. M. Disruption of imprinting caused by deletion of the *H19* gene region in mice. *Nature* **375**, 34–39 (1995).
- Lucifero, D., Mertineit, C., Clarke, H. J., Bestor, T. H. & Trasler, J. M. Methylation dynamics of imprinted genes in mouse germ cells. *Genomics* **79**, 530–538 (2002).
- Kono, T., Obata, Y., Yoshimizu, T., Nakahara, T. & Carroll, J. Epigenetic modifications during oocyte growth correlates with extended parthenogenetic development in the mouse. *Nature Genet.* **13**, 91–94 (1996).
- Obata, Y. *et al.* Disruption of primary imprinting during oocyte growth leads to the modified expression of imprinted genes during embryogenesis. *Development* **125**, 1553–1560 (1998).
- Feil, R., Walter, J., Allen, N. D. & Reik, W. Developmental control of allelic methylation in the imprinted mouse *Igf2* and *H19* genes. *Development* **120**, 2933–2943 (1994).
- Thorsvaldsen, J. L., Duran, K. L. & Bartolomei, M. S. Deletion of the *H19* differentially methylated domain results in loss of imprinted expression of *H19* and *Igf2*. *Genes Dev.* **12**, 3693–3702 (1998).
- Hark, A. T. *et al.* CTCF mediates methylation-sensitive enhancer-blocking activity at the *H19/Igf2* locus. *Nature* **405**, 486–489 (2000).
- Leighton, P. A., Saam, J. R., Ingram, R. S., Stewart, C. L. & Tilghman, S. M. An enhancer deletion affects both *H19* and *Igf2* expression. *Genes Dev.* **9**, 2079–2089 (1995).
- Ripoche, M. A., Kress, C., Poirier, F. & Dandolo, L. Deletion of the *H19* transcription unit reveals the existence of a putative imprinting control element. *Genes Dev.* **11**, 1596–1604 (1997).
- Kono, T., Sotomaru, Y., Katsuzawa, Y. & Dandolo, L. Mouse parthenogenetic embryos with monoallelic *H19* expression can develop to day 17.5 of gestation. *Dev. Biol.* **243**, 294–300 (2002).
- Wang, Z. Q., Fung, M. R., Barlow, D. P. & Wagner, E. F. Regulation of embryonic growth and lysosomal targeting by the imprinted *Igf2/Mpr* gene. *Nature* **372**, 464–467 (1994).
- Theiler, K. *The House Mouse: Atlas of Embryonic Development* 122–128 (Springer, Berlin, 1989).
- Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95**, 14863–14868 (1998).

23. Schmidt, J. V., Matteson, P. G., Jones, B. K., Guan, X. J. & Tilghman, S. M. The Dlk1 and Gtl2 genes are linked and reciprocally imprinted. *Genes Dev.* **14**, 1997–2002 (2000).
24. Kobayashi, S. *et al.* Mouse Peg9/Dlk1 and human PEG9/DLK1 are paternally expressed imprinted genes closely located to the maternally expressed imprinted genes: mouse Meg3/Gtl2 and human MEG3. *Genes Cells* **5**, 1029–1037 (2000).
25. Lin, S. P. *et al.* Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the Dlk1-Gtl2 imprinted cluster on mouse chromosome 12. *Nature Genet.* **35**, 97–102 (2003).
26. Georgiades, P., Watkins, M., Surani, M. A. & Ferguson-Smith, A. C. Parental origin-specific developmental defects in mice with uniparental disomy for chromosome 12. *Development* **127**, 4719–4728 (2000).
27. Georgiades, P., Watkins, M., Burton, G. J. & Ferguson-Smith, A. C. Roles for genomic imprinting and the zygotic genome in placental development. *Proc. Natl Acad. Sci. USA* **98**, 4522–4527 (2001).
28. Moore, T. & Haig, D. Genomic imprinting in mammalian development: a parental tug-of-war. *Trends Genet.* **7**, 45–49 (1991).
29. Reik, W. & Walter, J. Genomic imprinting: parental influence on the genome. *Nature Rev. Genet.* **2**, 21–32 (2001).
30. Murphy, S. K. & Jirtle, R. L. Imprinting evolution and the price of silence. *BioEssays* **25**, 577–588 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank A. Surani, Wellcome CRC Institute, J. Carroll, University College London and T. Moore, University College Cork, Ireland, for critical reading and discussions; O.-Y. Kwon, MacroGen, for microarray analysis; and T. Kumagai for technical assistance. This work was supported by grants from the Bio-oriented Technology Research Advancement Institution (BRAIN), Japan, and The Ministry of Education, Science, Culture and Sports of Japan.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.K. (tomohiro@nodai.ac.jp).

Intracellular gate opening in Shaker K⁺ channels defined by high-affinity metal bridges

Sarah M. Webster*, Donato del Camino*, John P. Dekker & Gary Yellen

Department of Neurobiology, Harvard Medical School, 220 Longwood Avenue, Boston, Massachusetts 02115, USA

* These authors contributed equally to this work

Voltage-gated potassium channels such as Shaker help to control electrical signalling in neurons by regulating the passage of K⁺ across cell membranes. Ion flow is controlled by a voltage-dependent gate at the intracellular side of the pore, formed by the crossing of four α -helices—the inner-pore helices. The prevailing model of gating is based on a comparison of the crystal structures of two bacterial channels—KcsA in a closed state and MthK in an open state—and proposes a hinge motion at a conserved glycine that splays the inner-pore helices wide open¹. We show here that two types of intersubunit metal bridge, involving cysteines placed near the bundle crossing, can occur simultaneously in the open state. These bridges provide constraints on the open Shaker channel structure, and on the degree of movement upon opening. We conclude that, unlike predictions from the structure of MthK, the inner-pore helices of Shaker probably maintain the KcsA-like bundle-crossing motif in the open state, with a bend in this region at the conserved proline motif (Pro-X-Pro) not found in the bacterial channels. A narrower opening of the bundle crossing in Shaker K⁺ channels may help to explain why Shaker has an approximately tenfold lower conductance than its bacterial relatives.

The recent elucidation of four K⁺ channel crystal structures has revealed the general architectural motifs in this large family of transmembrane proteins^{1–5}. Each of the tetrameric structures has a

central pore formed by two transmembrane helices and a selectivity filter structure from each of the four subunits. The sequence homology in the family of K⁺ channels is strongest in the region of the selectivity filter, a region that is crucial for allowing rapid and selective permeation of K⁺ ions⁶. Structurally, the selectivity filter is nearly identical in the channels solved thus far. However, the structures reveal significant differences in the conformation of the inner-pore helices in the region of the intracellular gate.

On the basis of differences in inner-helix configuration between two types of bacterial K⁺ channels, KcsA and MthK^{1,2,5}, MacKinnon and colleagues proposed a general model of K⁺ channel gating. In the apparently closed KcsA structure, the inner M2 helices are straight, and they cross in the lower part of the S6 to form an inverted teepee-like structure. Access from the intracellular surface to the pore is through a long, narrow, hydrophobic entrance. MthK, by contrast, was crystallized in conditions that favoured the open state of the channel. The inner helices are bent at Gly 83, which splay them apart to form a wide inner pore, 12 Å across at its narrowest point. These two structures are proposed to represent the closed and open states. In the gating model, the inner helices would swing around the conserved glycine to move from the narrow, closed KcsA pore to the wide, open MthK pore. This structure-based gating model is appealing in its simplicity, but it does not fit with the details of functional experiments done on voltage-gated K⁺ channels. Here we present a model of gating for a voltage-dependent K⁺ channel based on constraints established by the state dependence of metal bridging at two engineered binding sites.

A cysteine replacement at Shaker V476 creates a high-affinity Cd²⁺-binding site that stabilizes an open state of the channel⁷. Cd²⁺ binding prevents these channels from closing even at very negative voltages, apparently by forming a bridge between a cysteine in one subunit and a native histidine (H486) in a neighbouring subunit⁷ (Fig. 1). Replacement of the native histidine with tyrosine, threonine or valine abolished this high-affinity Cd²⁺ 'lock-open effect'. Other ligand combinations (Cys–Cys, His–Cys, His–His)

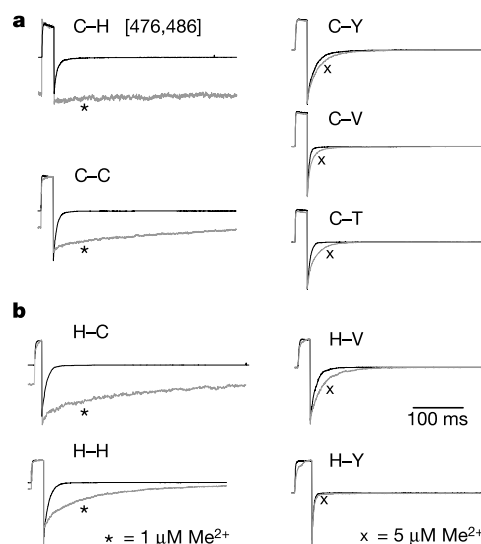


Figure 1 The lock-open effect requires metal binding residues at positions 476 and 486. Channel currents measured in control solutions (black traces) or in the presence of **a**, Cd²⁺ or **b**, Zn²⁺ (grey traces). Lock-open (prevention or delay of closure) occurs only for channels with two metal-binding ligands (left column), whereas channels with only one metal-binding ligand (right column) show only a slight slowing of the tails. Activating steps to +60 mV were followed by closure at –110 mV (C–C, C–H, C–V and C–T), –140 mV (H–C, H–H, H–V and C–Y) or –180 mV (H–Y). Without metal, maximum outward currents ranged from 0.7 to 2.7 nA; with metal, outward currents (normalized in the figure) ranged from 15 to 37% of the control because of inactivation.

could also stabilize the open state, although they were not quite as effective as the original Cys–His bridge. The Cd^{2+} lock-open effect requires specific metal-binding ligands, either cysteine or histidine, at both positions, 476 and 486, confirming that Cd^{2+} forms a bridge between these two positions.

This metal bridge can be used as a constraint on the open-state structure only if it does not distort the open state substantially. We used the binding of three open-channel blockers—tetraethylammonium (TEA), tetrabutylammonium (TBA) and the inactivation ball peptide (EBP)⁸—as one test for distortion of the open state, by comparing the affinity of the blockers for both the normal open state and the locked-open state. The blockers were applied by fast perfusion switching during a long open pulse either before or after the channel was locked open (Fig. 2a). For each of the blockers, the affinity was slightly better for locked-open channels than for channels in the absence of Cd^{2+} (fold increase: TEA, 2.6 ± 0.5 ; TBA, 2.2 ± 0.5 ; EBP, 4.0 ± 1.1). The structural changes produced by locking the channel open did not change blocker binding much (and actually slightly enhanced the binding). These changes correspond to a change in binding energy of only about 0.8–1.4 kJ, and are smaller than the ~ 2.3 kJ change in TEA affinity produced by removing four methylene groups in the T441S mutant⁹.

Many amino-acid substitutions in this region of the S6 affect the single-channel conductance by up to about fourfold^{10,11}, so it seems reasonable to use single-channel conductance as another test for distortion of the open state by the metal bridge. The single-channel current of V476C channels was not significantly different when the channel was locked open (Fig. 2b). The ratio of the single-channel current at 0 mV with and without 50 μM Cd^{2+} is 1.08 ± 0.12 ($n = 6$). As in the case with blocker affinity, this small change suggests that the bridge does not drastically alter the open state. A model of the open state should therefore be able to accommodate a metal bridge between V476C and H486 without much distortion.

At a second engineered site, V474C (at the position of the V in the

conserved Pro–Val–Pro), Cd^{2+} binds to multiple cysteines and blocks the current nearly irreversibly¹². Channels with only two cysteines at V474C bind Cd^{2+} reversibly, suggesting that the irreversible binding seen in a channel with four cysteines is due to the direct binding of three or four centrally facing cysteines. To achieve such tight binding, the sulphhydryl groups of the cysteine side chains must come into close proximity to coordinate a single Cd^{2+} ion jointly. A large change in the distance between these ligands upon channel opening, such as the change envisioned by the KcsA/MthK gating model, would predict state-dependent binding at this position, and Cd^{2+} binding should strongly influence the gating equilibrium.

To test this prediction, it was necessary to simplify the normally complex, multi-step process of gating. To accomplish this goal, these studies were conducted in a mutant background that separates the voltage range of channel opening from that of the multi-step voltage sensor movement. This background contains three conservative mutations of uncharged residues in the S4 helix: V369I, I372L and S376T (ILT)^{13–15}. We introduced the V474C mutation into the ILT background (ILT 474C) and confirmed that it maintained the essential hallmarks of the ILT two-state model: voltage separation of the major charge movement and channel opening, a reduced slope and positive-shifted midpoint of the gating curve, and single exponential opening kinetics after a short delay (data not shown).

Because V474C channels do not conduct current when Cd^{2+} is bound, the gating equilibrium of Cd^{2+} -bound channels cannot be measured directly. However, the release of Cd^{2+} in the presence of a dithiol reagent, 2,3-dimercaptopropanesulphonate (DMPS), is steeply voltage dependent (in both the ILT and WT¹² backgrounds) in a way that suggests that Cd^{2+} -bound channels still have voltage-dependent opening, with release occurring only from the open state. Therefore, the gating of Cd^{2+} -bound channels was inferred from the voltage dependence of recovery. In 0.5 mM DMPS, the recovery rates of ILT 474C had a voltage dependence similar to that of gating

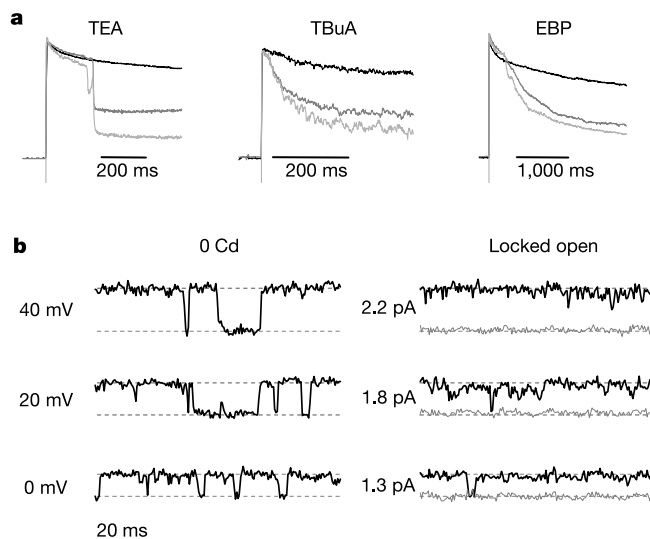


Figure 2 The blocker affinity and single-channel conductance of the locked-open state are similar to those of the normal open state. **a**, Open channel blockers applied in the middle of a step to 0 mV give nearly the same block whether the channel is open (dark grey) or locked open (with prior exposure of 10 μM Cd^{2+} , light grey; no blocker, black). Blocker concentrations: TEA, 1 mM; TBA, 25 μM ; EBP, 0.4 μM . Control currents ranged from 0.2 to 1 nA; with metal, outward currents (normalized in the figure) ranged from 42% to 80% of control because of inactivation. **b**, Single-channel recordings from 476C channels in 0 μM Cd^{2+} (left) and 50 μM Cd^{2+} (right). Dotted lines show open and closed levels; the light grey trace shows the closed level in a subsequent pulse where the channel had inactivated.

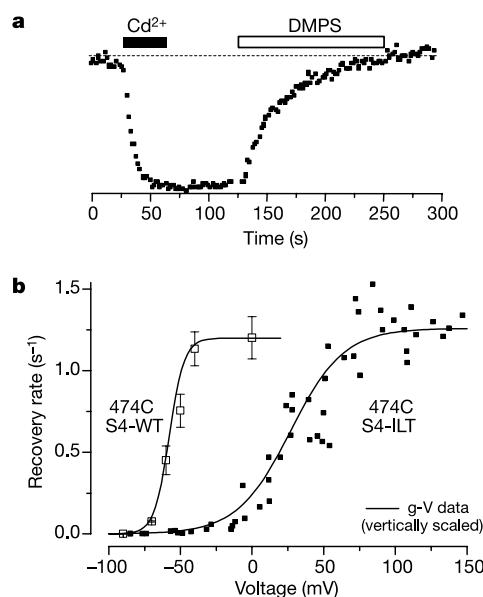


Figure 3 The recovery of ILT V474C channels from Cd^{2+} blockade suggests that cadmium has little effect on the open/closed equilibrium of bound channels. **a**, Recovery was measured by pulsing (+120 mV for 40 ms from -90 mV) in the presence of 0.5 mM DMPS (squares, current for each pulse; dotted line, 700 pA). **b**, The recovery rate of ILT V474C from Cd^{2+} block as a function of voltage with 0.5 mM DMPS (filled squares). The Cd^{2+} recovery rate for WT 474C from ref. 12 (open squares $\pm$ s.e.m.) and the corresponding g-Vs (line with vertical scaling to the recovery rates) are plotted for comparison.

in ILT V474C (Fig. 3). This agrees with the earlier result in a wild-type S4 background¹² (Fig. 3b), which has more complicated gating. Although the ILT mutation shifts the voltage dependence by over 90 mV and makes it much shallower, the voltage dependence of Cd^{2+} -free gating (the g -V) agrees with that for Cd^{2+} recovery in both cases. This, and the agreement between the maximal rate of recovery in the two mutants, makes a strong case for recovery occurring only from the open state, and that the open-closed equilibrium is not altered by Cd^{2+} binding. The relative lack of effect on the gating equilibrium suggests that Cd^{2+} is bound almost equally well in the open and closed states, suggesting that there is little change in the positions of the side chains of V474C with opening.

In two separate sets of experiments, we have shown that the two different metal bridges, V474C and V476C–H486, each occur with high affinity in the open state. We reasoned that if neither metal binding event distorted the open-state structure, then Cd^{2+} should bind both sites simultaneously in the open state. To test this prediction, we measured the ability of the double mutant V474C–V476C to bind Cd^{2+} at V474C after being locked open at V476C–H486 (Fig. 4). The channels were first locked open by a 380 ms application of Cd^{2+} to the intracellular surface of the channel, with a pore blocker (hexyltriethylammonium; C_6 -TEA) present to protect V474C from modification. This treatment produced a clear lock-open effect, as shown by the sustained inward current during the negative test pulses. When the locked-open channels were exposed a second time to Cd^{2+} , there was rapid block at V474C (the rate constant was $3.9 \pm 0.2 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$, $n = 3$, see Methods). This on rate for the block at V474C was about sevenfold slower than in a non-locked-open channel¹⁶, but given the precise distance and geometry constraints required for Cd^{2+} binding to multiple cysteines, it does not seem surprising that even a small structural change produces a change in the on rate. We know that the equilibrium binding of Cd^{2+} to 474C remains very strong—essentially irreversible—even when the binding occurs with the channels locked open. Despite this change in kinetics, it appears that bridges for both the lock-open effect and for blockade can be accommodated simultaneously by the open state.

Both of the metal-binding bridges that we observe in the open state of Shaker are incompatible with the MthK-derived open-state model. By comparison with known structures with Cys– Cd^{2+} –Cys bridges (Cys– C_β to Cys– C_β of 4.0–6.9 Å, for example, in PDB 1D66),

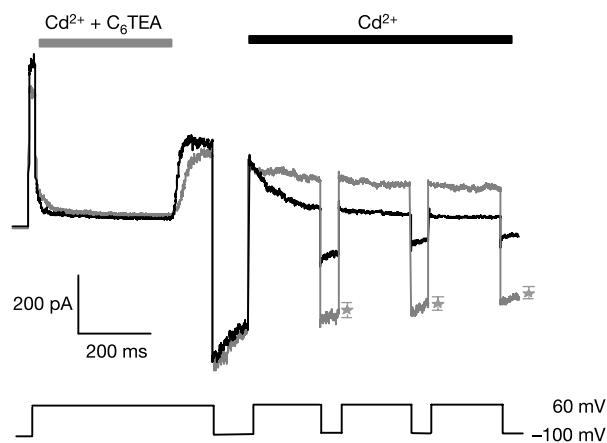


Figure 4 Cadmium ions can simultaneously bind V474C and bridge V476C–H486. 2 mM C_6 -TEA was added to protect 474C while 40 μM Cd^{2+} is applied to lock the channels open (grey bar). Sustained inward current at negative pulses shows the extent of the lock-open effect (control, grey trace; stars, the mean current in control ($\pm$ s.e.m., $n = 3$); black line, voltage protocol). The decay of current at negative voltages during a second application of Cd^{2+} (black bar, black trace) monitors binding to V474C.

the β carbons in MthK that are equivalent to the 474C– Cd^{2+} –474C bridges are 5–12 Å too far apart (Fig. 5a, b, right). A similar analysis for the intersubunit 476C– Cd^{2+} –H486 positions estimates that these positions in MthK are ~ 10 –15 Å more distant than required.

Are these interactions actually occurring in the open state? Our use of the Shaker ILT mutation makes channel gating much simpler, but it does not completely eliminate the possibility of unobserved intermediate states. It remains possible that after the weakly voltage-dependent ‘opening’ transition in ILT, the channel undergoes a further rapid, voltage-independent transition between a narrow ‘snug’ (nonconductive) state and a wide, open state. In this alternative model, the snug state would permit the tight binding of cadmium and probably also the tight binding of quaternary ammonium blockers¹⁷. Although the open state in this model might be wide, its structure would be quite different from that of MthK to be consistent with the lock-open bridge. Our recovery data would be compatible with such a model only if a very specific coincidence (for both wild-type S4 and ILT channels) stabilizes the Cd^{2+} -bound snug state by an amount that almost exactly offsets the stability normally provided by coupling to the open state. A more serious challenge to this model is the observed rate of Cd^{2+} binding to 474C in the locked-open channels (Fig. 4). To explain this, we would need to suppose that the lock-open bridge decreases the occupancy of the snug state by no more than about sevenfold, which seems incompatible with the fact that the bridge reduces the closing rate by at least 1,000-fold (the dwell time of Cd^{2+} on the lock-open site is nearly a second, even at very negative voltages). Moreover, if the snug state provides the best fit for blockers like TEA and TBuA¹⁷, then locking the channels in the open state ought to reduce the affinity, contrary to our observation (Fig. 2a). Taken together, the most straightforward interpretation of this body of functional data is that both metal bridges occur in the open state, and thus the structure of the Shaker inner helix in the open (and closed¹⁵) states is significantly different from that of the distant bacterial relatives KcsA and MthK (although no more different than is common within families of structurally related proteins).

The two bridges defined here provide strong constraints on the structure of the open state. Our proposed ‘bent S6’ model (centre, Fig. 5b) was derived by modifying the KcsA backbone to account for functional discrepancies with the KcsA structure^{16,18}. The voltage-gated K^+ channel family has a highly conserved Pro–X–Pro motif in the middle of the S6, just at the bundle crossing in KcsA, which could serve as a helix-breaking sequence^{19,20}. In our model, a kink in the KcsA helices was introduced at the Pro–Val–Pro (473–475) motif to bring V476C and H486 within distances appropriate for metal binding. This bent S6 model places both Cd^{2+} sites in acceptable binding positions simultaneously. Although we do not think that the open Shaker channel is likely to be quite so narrow (4–7 Å) as to allow optimum placement of the 474 cysteines for the binding of Cd^{2+} , a slightly larger opening (for example, ~ 8 –9 Å) would permit hydrated K^+ ions (cross-section $\sim 6 \times 6$ Å)²¹ and large quaternary ammonium blockers (such as TBuA or tetrahexylammonium; cross-sections of $\sim 4.5 \times 8$ Å) to enter the cavity^{17,22}, and would allow Cd^{2+} to bind with only minor flexibility (see ref. 23). The ability to assume the very narrow Cd^{2+} -binding position with equal ease from the open and closed states argues against a large motion at this position and for a constricted entrance to the cavity even in the open state.

The results indicate that at position V474 the intracellular entrance to the pore cavity of Shaker K^+ channels can be narrow in both states. The constriction of ion flow in the closed state is below this point, probably at position V478 (refs 16, 24). How, then, does the channel gate? Use of the conserved Pro–X–Pro as a flexible hinge or swivel could keep the upper S6 (and position 474) relatively still, but allow the lower half of the S6 to move, opening the pore by swivelling the blocking residues away from the central axis (Fig. 5). This is a different kind of movement than the large outward swing of

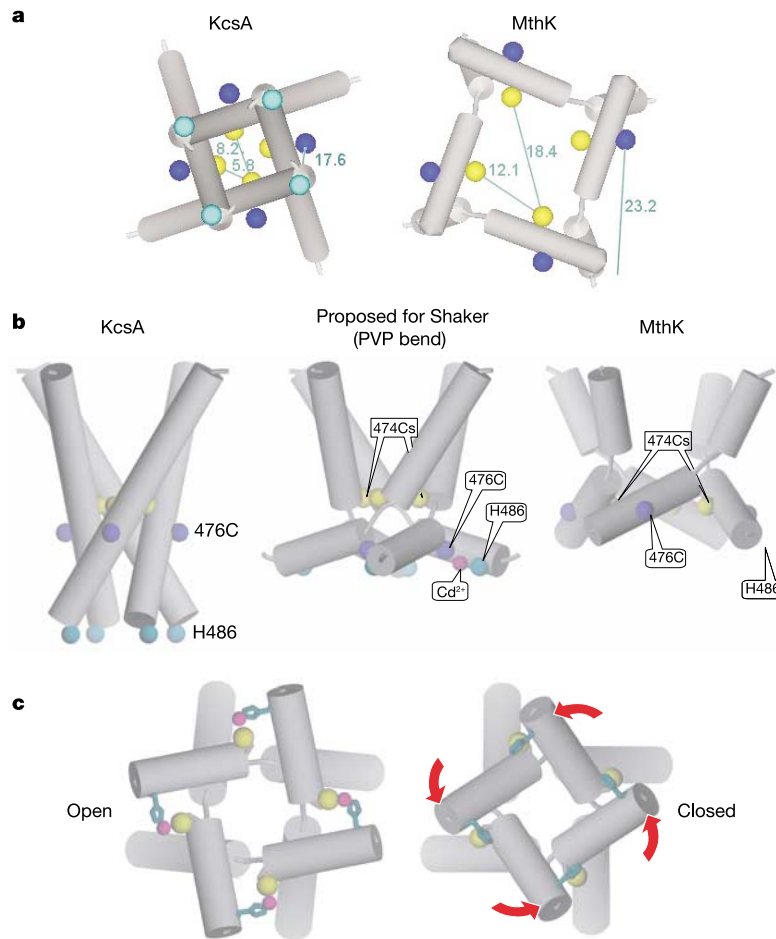


Figure 5 A comparison of the models of potassium channel gating. **a**, The distances between the positions equivalent to V474C (yellow) or between V476C (blue) and H486 (cyan) in the KcsA (left) or MthK (right) structures shown from the intracellular surface (position of the MthK 486-equivalent was estimated from a linear extension of the helix).

b, The proposed Shaker pore structure (centre) is based on KcsA (left) but places a bend in the S6 at the conserved Pro-X-Pro motif (side view, colours as in **a** with Cd^{2+} in magenta). **c**, An intracellular view of a gating model based on the two Cd^{2+} bridges and previous constraints on the closed state¹⁶.

helices that would result from a motion around the proposed hinge at the conserved glycine in the KcsA–MthK model of gating. In a combination of the two models, a similar opening motion might occur by swivelling at the glycine hinge with a fixed bend at the Pro-Val-Pro while maintaining the proximity of neighbouring V474C residues (that is, still pivoting around the Pro-Val-Pro).

Does the narrow opening of Shaker limit the maximum conductance of the channel? It has been proposed that the lower end of the S6 of Shaker contributes to a barrier for ion flow^{10,11}. The MthK single-channel conductance⁵ is ~ 220 pS, whereas Shaker conductance is about tenfold smaller, ~ 24 pS, in similar conditions²⁵. The negative charges found in MthK and another large conductance channel, BK, account for only a small part of the difference in the conductance, only about twofold^{26,27}. To account for the large difference in current flow, we suggest that a narrow opening in the class of smaller conductance channels such as Shaker places a diffusional barrier to ion conduction in the inner pore. Changing the geometry of the inner pore could modulate this conduction barrier, allowing for the tuning of single-channel conductance while maintaining strict ion preference in the selectivity filter. □

Methods

Molecular biology

Mutations were introduced by polymerase chain reaction (PCR) mutagenesis into our modified 'wild-type' cDNA¹⁵, Shaker H4 $\Delta 6$ -46 C301S C308S T449V. The ILT mutations (V369I, I372L and S376T) were transferred to our background by PCR generation of a

fragment that was subcloned into the *Xba*I/*Bgl*II sites. Because the V476H mutant in a wild-type background opens at very negative voltages, we placed the mutation in an S4-LT (I372L, S376T) background¹⁴ to shift gating into a more accessible voltage range. All mutations were confirmed by sequencing.

Recording and solutions

Channels were transiently expressed in HEK 293 cells by electroporation as described previously¹⁵; recordings were performed 1–3 days after transfection. All experiments were conducted in inside-out patches pulled from identified transfected cells. The methods for rapid perfusion switches and electrophysiological recordings were as previously described¹², except in Fig. 4. In Fig. 4, a piezoelectric-driven solution-switching mechanism was used with a four-barrelled perfusion pipette²⁸. Intracellular solutions contained (in mM) 160 KCl, 10 HEPES, 1 EGTA and 0.5 MgCl_2 (bath). Extracellular solutions contained (in mM) 60 KCl, 100 NaCl, 3 CaCl_2 , 1 MgCl_2 and 10 HEPES, except in the experiments of Fig. 2b and Fig. 3, which contained 10 KCl and 150 NaCl.

When Cd^{2+} or Zn^{2+} were included, the EGTA was left out of the solution. The experimental cadmium and zinc concentrations are reported as the total concentration of metal. However, we used the chloride-buffer-corrected free concentration of Cd^{2+} to calculate the on rate constant in Fig. 4 (in the presence of 160 mM chloride ion, the calculated ratio of free $[\text{Cd}^{2+}]$ to total $[\text{Cd}^{2+}]$ is 0.13). For comparison, the previously reported on rate of Cd^{2+} to the single-mutant V474C channels¹⁶ is $2.9 \pm 0.4 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ ($n = 4$) when corrected for buffering in the same way.

In the ILT mutants, there was evidence of a voltage-dependent block at positive voltages, which was alleviated by the removal of MgCl_2 from the intracellular solutions; the ILT experiments shown were carried out in the absence of internal MgCl_2 .

The enhanced ball peptide we used in Fig. 2 has the sequence of the first 20 amino acids of Shaker with two mutations, E12K and E13K, to increase the positive charge⁶. DMPS stock solutions in water (50 mM) were made daily and kept on ice. Dilutions into intracellular solution were made just before each experiment.

Data analysis

Because channels continue to undergo slow inactivation when they are locked open, even

at very negative potentials, the lock-open currents were normalized to the current level in the control case for comparison. To estimate the $K_{1/2}$ of the blockers, the amount of block was estimated by isochronal comparison of the scaled traces.

In the single-channel recordings, the leak current and capacitive transients were subtracted with an idealized fit to a blank trace. Traces were filtered at 1 kHz, sampled at 2 kHz.

The rates of recovery of Shaker V474C from block by Cd^{2+} by DMPS were measured as the single exponential fit of the current as a function of the cumulative exposure at the test voltage. The test pulses were adjusted in length so that the time constant of recovery would be equal to at least two pulse durations. These single exponential rates were plotted versus voltage. Because of the large variation in the g-V midpoint for each experiment, the voltage of the recovery was corrected for the g-V midpoint of the individual experiment¹³: $V_{\text{corrected}} = V + (V_{\text{mid}} - V_{\text{mid}})$. The average Boltzmann fit of the g-Vs is plotted.

Received 20 November 2003; accepted 4 March 2004; doi:10.1038/nature02468.

- Jiang, Y. *et al.* The open pore conformation of potassium channels. *Nature* **417**, 523–526 (2002).
- Doyle, D. A. *et al.* The structure of the potassium channel: molecular basis of K^+ conduction and selectivity. *Science* **280**, 69–77 (1998).
- Kuo, A. *et al.* Crystal structure of the potassium channel KirBac1.1 in the closed state. *Science* **300**, 1922–1926 (2003).
- Jiang, Y. *et al.* X-ray structure of a voltage-dependent K^+ channel. *Nature* **423**, 33–41 (2003).
- Jiang, Y. *et al.* Crystal structure and mechanism of a calcium-gated potassium channel. *Nature* **417**, 515–522 (2002).
- Heginbotham, L., Lu, Z., Abramson, T. & MacKinnon, R. Mutations in the K^+ channel signature sequence. *Biophys. J.* **66**, 1061–1067 (1994).
- Holmgren, M., Shin, K. S. & Yellen, G. The activation gate of a voltage-gated K^+ channel can be trapped in the open state by an intersubunit metal bridge. *Neuron* **21**, 617–621 (1998).
- Zagotta, W. N., Hoshi, T. & Aldrich, R. W. Restoration of inactivation in mutants of Shaker potassium channels by a peptide derived from ShB. *Science* **250**, 568–571 (1990).
- Yellen, G., Jurman, M. E., Abramson, T. & MacKinnon, R. Mutations affecting internal TEA blockade identify the probable pore-forming region of a K^+ channel. *Science* **251**, 939–942 (1991).
- Ding, S. & Horn, R. Tail end of the S6 segment: Role in permeation in Shaker potassium channels. *J. Gen. Physiol.* **120**, 87–97 (2002).
- Lopez, G. A., Jan, Y. N. & Jan, L. Y. Evidence that the S6 segment of the Shaker voltage-gated K^+ channel comprises part of the pore. *Nature* **367**, 179–182 (1994).
- Liu, Y., Holmgren, M., Jurman, M. E. & Yellen, G. Gated access to the pore of a voltage-dependent K^+ channel. *Neuron* **19**, 175–184 (1997).
- Smith-Maxwell, C. J., Ledwell, J. L. & Aldrich, R. W. Role of the S4 in cooperativity of voltage-dependent potassium channel activation. *J. Gen. Physiol.* **111**, 399–420 (1998).
- Smith-Maxwell, C. J., Ledwell, J. L. & Aldrich, R. W. Uncharged S4 residues and cooperativity in voltage-dependent potassium channel activation. *J. Gen. Physiol.* **111**, 421–439 (1998).
- Ledwell, J. L. & Aldrich, R. W. Mutations in the S4 region isolate the final voltage-dependent cooperative step in potassium channel activation. *J. Gen. Physiol.* **113**, 389–414 (1999).
- del Camino, D. & Yellen, G. Tight steric closure at the intracellular activation gate of a voltage-gated K^+ channel. *Neuron* **32**, 649–656 (2001).
- Zhou, M., Morais-Cabral, J. H., Mann, S. & MacKinnon, R. Potassium channel receptor site for the inactivation gate and quaternary amine inhibitors. *Nature* **411**, 657–661 (2001).
- del Camino, D., Holmgren, M., Liu, Y. & Yellen, G. Blocker protection in the pore of a voltage-gated K^+ channel and its structural implications. *Nature* **403**, 321–325 (2000).
- Tieleman, D. P., Shrivastava, I. H., Ulmschneider, M. R. & Sansom, M. S. Proline-induced hinges in transmembrane helices: possible roles in ion channel gating. *Proteins* **44**, 63–72 (2001).
- Bright, J. N., Shrivastava, I. H., Cordes, F. S. & Sansom, M. S. Conformational dynamics of helix S6 from Shaker potassium channel: simulation studies. *Biopolymers* **64**, 303–313 (2002).
- Zhou, Y., Morais-Cabral, J. H., Kaufman, A. & MacKinnon, R. Chemistry of ion coordination and hydration revealed by a K^+ channel-Fab complex at 2.0 Å resolution. *Nature* **414**, 43–48 (2001).
- Holmgren, M., Smith, P. L. & Yellen, G. Trapping of organic blockers by closing of voltage-dependent K^+ channels: evidence for a trap door mechanism of activation gating. *J. Gen. Physiol.* **109**, 527–535 (1997).
- Loussouarn, G., Phillips, L. R., Masia, R., Rose, T. & Nichols, C. G. Flexibility of the Kir6.2 inward rectifier K^+ channel pore. *Proc. Natl Acad. Sci. USA* **98**, 4227–4232 (2001).
- Hackos, D. H., Chang, T. H. & Swartz, K. J. Scanning the intracellular S6 activation gate in the Shaker K^+ channel. *J. Gen. Physiol.* **119**, 521–532 (2002).
- Isacoff, E. Y., Jan, Y. N. & Jan, L. Y. Putative receptor for the cytoplasmic inactivation gate in the Shaker K^+ channel. *Nature* **353**, 86–90 (1991).
- Nimigeon, C. M., Chappie, J. S. & Miller, C. Electrostatic tuning of ion conductance in potassium channels. *Biochemistry* **42**, 9263–9268 (2003).
- Brelidze, T. I., Niu, X. & Magleby, K. L. A ring of eight conserved negatively charged amino acids doubles the conductance of BK channels and prevents inward rectification. *Proc. Natl Acad. Sci. USA* **100**, 9017–9022 (2003).
- Forman, S. A. A hydrophobic photolabel inhibits nicotinic acetylcholine receptors via open-channel block following a slow step. *Biochemistry* **38**, 14559–14564 (1999).

Acknowledgements We thank M. Kanevsky for characterizing the charge movement for ILT 474C and S. Forman for allowing us to use his perfusion setup. We are also grateful to B. Bean and to the members of the Yellen laboratory for helpful discussions, and to T. Abramson for her expert help with transfections. This work was supported by a grant from the NIH/NINDS to G.Y.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.Y. (gary_yellen@hms.harvard.edu).

Programmed population control by cell-cell communication and regulated killing

Lingchong You¹, Robert Sidney Cox III², Ron Weiss³ & Frances H. Arnold¹

¹Division of Chemistry and Chemical Engineering and ²Division of Biology, California Institute of Technology, Pasadena, California 91125, USA

³Departments of Electrical Engineering and Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA

De novo engineering of gene circuits inside cells is extremely difficult^{1–9}, and efforts to realize predictable and robust performance must deal with noise in gene expression and variation in phenotypes between cells^{10–12}. Here we demonstrate that by coupling gene expression to cell survival and death using cell-cell communication, we can programme the dynamics of a population despite variability in the behaviour of individual cells. Specifically, we have built and characterized a ‘population control’ circuit that autonomously regulates the density of an *Escherichia coli* population. The cell density is broadcasted and detected by elements from a bacterial quorum-sensing system^{13,14}, which in turn regulate the death rate. As predicted by a simple mathematical model, the circuit can set a stable steady state in terms of cell density and gene expression that is easily tunable by varying the stability of the cell-cell communication signal. This circuit incorporates a mechanism for programmed death in response to changes in the environment, and allows us to probe the design principles of its more complex natural counterparts.

Our circuit (Fig. 1a) programmes a bacterial population to maintain a cell density that is lower than the limits imposed by the environment (for example, by nutrient supply). The LuxI protein of the well-characterized LuxI/LuxR system from the marine bacterium *Vibrio fischeri*^{13,14} synthesizes a small, diffusible acyl-homoserine lactone (AHL) signalling molecule. The AHL accumulates in the experimental medium and inside the cells as the cell density increases. At sufficiently high concentrations, it binds and activates the LuxR transcriptional regulator, which in turn induces the expression of a killer gene (*E*) under the control of a *luxI* promoter (p_{luxI})¹⁵. Sufficiently high levels of the killer protein cause cell death.

We implemented the circuit using a two-plasmid system (Fig. 1b), where pLuxRI2 expresses LuxI and LuxR upon induction by isopropyl-β-D-thiogalactopyranoside (IPTG), and *pluxCcdB3* responds to activated LuxR (at sufficiently high cell density) and causes cell death. The *lacZα-ccdB* killer gene codes for a fusion protein of LacZα and CcdB. The LacZα portion of the fusion protein retains the ability to complement LacZΔM15 in appropriate cell strains (for example, Top10F' cells), allowing the measurement of fusion protein levels using a LacZ assay (see Methods). The CcdB portion retains the toxicity of native CcdB, which kills susceptible cells by poisoning the DNA gyrase complex¹⁶.

A simple mathematical model predicted that the system would reach a stable cell density for all realistic parameter values (see Methods), although it might go through damped oscillations while approaching the steady state. Experiments confirmed our predictions. Figure 2a shows the growth of Top10F' cells containing the population-control circuit at pH 7.0. As anticipated, the uninduced culture (circuit OFF) grew exponentially and reached the stationary phase upon nutrient exhaustion. The induced culture (circuit ON) grew almost identically, until its density reached a threshold (at 7 h). It then deviated sharply from the OFF culture and briefly went through a damped oscillation (between 7 h and 24 h) of at least one cycle before reaching a steady-state density about ten times lower than that of the OFF culture. The measured peak density (at 10 h)

was twofold higher than the measured floor (12.5 h), a difference significantly greater than measurement variations. The steady state was maintained for more than 30 h: the variation in cell density between 28 h and 62 h was <5%, which was smaller than typical measurement variations of individual data points. The transient dynamics of circuit-ON growth was captured well in simulation, where the AHL degradation rate constant (d_A) was adjusted for the simulation to match the experiment in the steady-state circuit-ON cell density. Intracellular levels of LacZ α -CcdB for the ON culture, measured in terms of LacZ activity, reached a steady state after an overshoot (Fig. 2b), again predicted by the simulation. The basal level of LacZ α -CcdB expression in the OFF culture was negligible for all time points.

The growth curves and the LacZ α -CcdB time courses illustrate the tight yet dynamic coupling between population dynamics and intracellular expression of the LacZ α -CcdB killer protein. A level of the killer protein lower than steady state allows the population to grow; conversely, its excessive production decreases cell density. After some delay, the decline in cell density leads to a decrease in AHL concentration, which in turn leads to reduced levels of the killer protein, allowing the population to recover. Continuous production and degradation (or death) of each circuit component are essential for the observed homeostasis, and they are closely coupled. Any perturbation that decouples or overrides these processes will disrupt circuit function. For example, a circuit without *luxI* (thus lacking a communication link) could not control growth (data not shown). Also, we observed that 200 nM exogenous AHL, which was not toxic to cells without the circuit or with the circuit OFF, completely prevented growth with the circuit ON (data not shown). This is expected, because a high level of AHL would activate LuxR and lead to overproduction of the killer protein. This observation also confirms that the killer protein production rate was limited by AHL synthesis in circuit-ON growth, a key assumption in our mathematical model.

Circuit function could also be delicately manipulated in a predictable fashion. Our model predicted that the steady-state cell density would increase nearly proportionally with the AHL degradation rate constant (see Methods). Thus, AHL serves as an external 'dial' to operate the circuit: AHL degradation affects cell-cell communication, and rapid AHL breakdown can disrupt it completely. Degradation of AHL is facilitated by hydrolytic enzymes^{17,18} or by increasing the medium pH¹⁹. Confirming the model prediction, a moderate increase in the medium pH (6.2 to

7.8) significantly increased (~fourfold) steady-state cell densities of the ON cultures, but caused only minor changes in those of the OFF cultures (Fig. 3a–e, Table 1). For each pH value, circuit-ON populations reached a steady state after 28 h, and the variation in cell density afterwards was smaller than typical measurement variations. Similar to the pH 7.0 case, the ON cultures grew almost identically to the OFF cultures at low cell density, but deviated from the latter at high density. Again, simulations captured the experimental behaviour (Fig. 3a–d), with adjustment only of the AHL degradation rate constant (Table 1).

Our model predicted that, unlike cell density, intracellular levels of the killer protein would remain roughly constant as pH varied. At steady state, $E_s = k/d (1 - N_s/N_m) \approx k/d$, the ratio of the growth and killing rate constants (assuming that the circuit-ON cell density is far below the carrying capacity; that is, $N_s/N_m \ll 1$). Thus if pH did not significantly affect the growth or killing rate constants, the killer protein would reach a concentration of k/d , independently of pH. Experimental results showed that LacZ α -CcdB for different pHs reached similar levels after about 28 h (Fig. 3f–i, Table 1). Nevertheless, the pH changes affected other measured parameters, including k , N_m and more notably N_s (Table 1). To account for the effects of these changes on killer protein expression, we normalized LacZ activity with respect to $k(1 - N_s/N_m)$. According to the model, the normalized LacZ activity should remain constant if pH does not affect the killing rate constant (because $E_s/(k(1 - N_s/N_m)) = 1/d$). Supporting the prediction, normalized LacZ activities were nearly identical for different pHs (Fig. 3j). Except for pH 8.05, these values were statistically indistinguishable ($P = 0.26$). The slightly, but significantly, higher ($P = 8.6 \times 10^{-4}$) normalized LacZ activity at pH 8.05 suggests that the killer protein was less toxic (smaller d).

Construction of *de novo* gene circuits can be used to decode 'design principles' of biological systems^{20–22}. Studies have demonstrated the feasibility of building gene circuits that lead to stable¹, bi-stable^{2–4} or oscillatory^{3,5} gene expression, act as a digital logic inverter^{6,7} or a metabolic controller⁸, or provide better-regulated gene expression systems⁹. They have also revealed major hurdles to achieving reliable circuit behaviour, such as noise in cellular processes and cell-to-cell variation^{10–12}. Here we addressed these issues by using cell-cell communication to coordinate behaviour across the population. By coupling gene expression with population dynamics, cell-cell communication integrates the entire population as an essential component of the population-control circuit. This

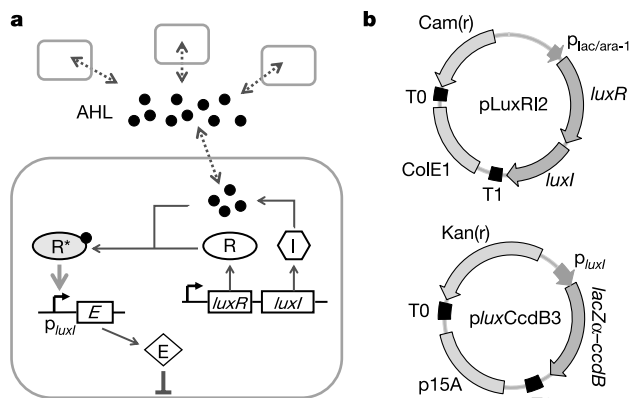


Figure 1 A population-control circuit programmes population dynamics by broadcasting, sensing and regulating the cell density using cell-cell communication and negative feedback. **a**, Schematic diagram of the circuit. *E* is a 'killer' gene. *I*, *R* and *R** represent LuxI, LuxR and active LuxR, respectively. Filled circles represent AHL. **b**, Experimental implementation with two plasmids, pLuxRI2 and pluxCcdB3. *luxI* and *luxR* are under the control of a synthetic promoter $P_{lac/ara-1}$ ³², and the killer gene (*lacZ α -ccdB*) is under the control of P_{luxI} ¹⁵. T0 and T1 are transcription terminators. See text for details.

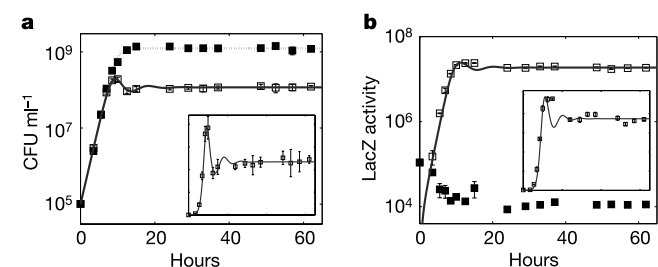


Figure 2 Experimentally measured growth curves (**a**) and corresponding levels of LacZ α -CcdB (**b**) of Top10F' cells with the population-control circuit OFF (filled squares) and ON (open squares) for pH 7.0. Model predictions are shown in solid (ON) and dotted (OFF) lines, except for the OFF case of LacZ activity, where the killer protein concentration is always zero in the model. The simulated LacZ activity is obtained by multiplying the simulated killer protein concentration (in nM) by a constant factor so that the experiment and simulation are at the same scale. Insets show the growth curves and the LacZ activity in linear scale for the ON case. The similar growth of two cultures at low cell density is an intrinsic feature of the circuit and is not caused by a lag in circuit activation. When the ON culture at steady state was diluted into fresh medium with and without the inducer, the resulting cultures again grew similarly at low density but deviated at high density.

Table 1 Effects of pH on circuit parameters

Medium pH	Steady-state culture pH*	$k\uparrow$ (h ⁻¹)	N_m per 10 ⁹ ‡ (CFU ml ⁻¹)	N_s per 10 ⁷ § (CFU ml ⁻¹)	d_{All} (h ⁻¹)	LacZ activity/10 ⁷ ¶ (fluorescence per (ml × OD 600))
6.2	6.45	0.885	1.25 ± 0.06	4.86 ± 0.02	0.274	1.94 ± 0.12
6.6	6.76	0.928	1.17 ± 0.05	5.59 ± 0.03	0.304	2.02 ± 0.17
7.0	7.18	0.970	1.24 ± 0.10	11.7 ± 0.6	0.639	1.87 ± 0.09
7.4	7.53	0.897	1.16 ± 0.10	13.1 ± 0.6	0.791	1.79 ± 0.16
7.8	8.05	0.936	1.20 ± 0.07	19.5 ± 1.3	1.19	2.00 ± 0.06

*Measured at about 50 h after growth initiation in ON cultures.

†Obtained by fitting the exponential phase of growth curves of OFF cultures.

‡Average of the stationary-phase cell density of OFF cultures between 28 h and 62 h.

§Average of the circuit-ON cell density between 28 h and 62 h.

||Obtained by solving equation $N_s = \frac{N_m d_{All} d_{EK}}{N_m v_A k_E d + d_{All} d_{EK}}$ with d_{All} as the only unknown.

¶Average of LacZ activity of ON cultures between 28 h and 62 h.

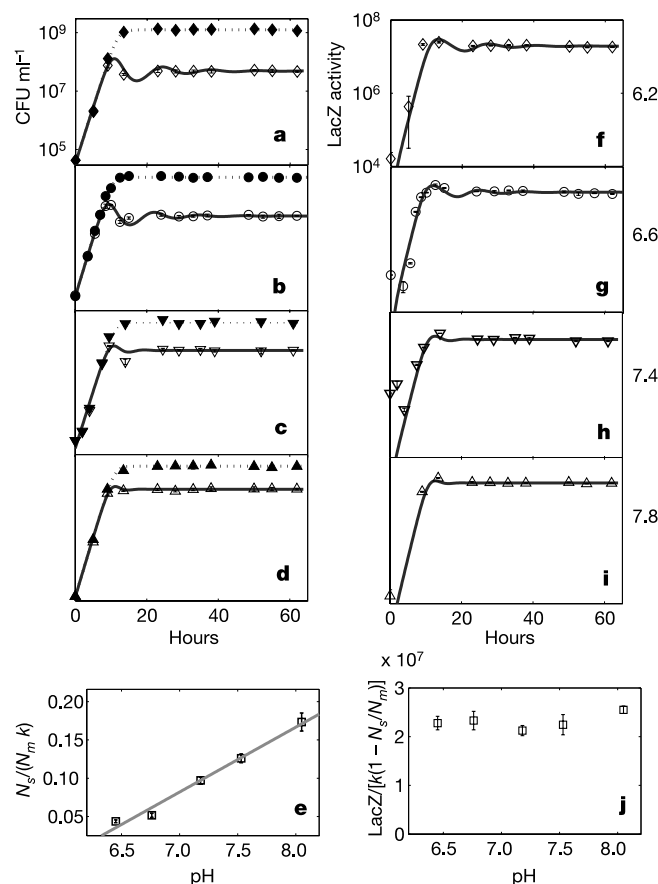


Figure 3 Effects of pH on circuit behaviour. **a–d**, Cell growth with the circuit OFF (filled symbols) and ON (open symbols). **e**, Dependence of $N_s/(N_m \times k)$ on pH. **f–i**, Time courses of LacZ activity with the circuit ON. **j**, Dependence of $\frac{\text{LacZ activity}}{k(1 - N_s/N_m)}$ on pH. Panels **a–d** have the same scale in both x and y axes, as do the panels **f–i**. Simulated growth curves (ON, solid line; OFF, dotted line) and killer protein time courses (ON, solid line) are shown in **a–d** and **f–i**. The killer protein concentration for the OFF cases is always zero in the model. Simulated LacZ activity in **f–i** is obtained by multiplying the simulated killer protein concentration by a constant factor so that the experiment and simulation are at the same scale in each panel. In **e** and **j**, steady-state pH values (Table 1) are used along the x axes. In **e**, N_s is normalized with respect to N_m and k to account for minor variations in these variables (Table 1). The dependence of $N_s/(N_m \times k)$ on pH is nearly linear ($R^2 = 0.98$).

coupling enables the circuit to function reliably at the population level by exploiting cell heterogeneity in terms of their size, age, plasmid copy number, gene expression and response to the killer protein. Such phenotypic variations, which interfere with the functioning of circuits lacking a communication mechanism, are actually required for the population-control circuit to work. The

outcome of circuit function is binary for any given cell: it lives or dies (judged by the ability to form a colony). If all cells had the same phenotype, the circuit would fail to achieve a stable cell density because the population would become extinct once the killer protein concentration reached a critical threshold. And because of the coupling, the circuit can function only at the population level. This is supported by experimental data: under each set of conditions, growth of the circuit-ON culture only deviated from that of the OFF culture when cell density was sufficiently high (Figs 2 and 3). If the circuit had functioned inside each cell autonomously, growth of the two cultures would have deviated from the beginning.

Similar to multicellular organisms, bacteria possess sophisticated suicide machinery that is triggered by stress and starvation or by 'addiction modules' during post-segregational killing^{16,23,24}. Natural systems, such as the signalling network that dictates the lysis of a subpopulation of *Streptococcus pneumoniae* at high cell density, employ a mechanism similar to our circuit of regulated killing coordinated by quorum sensing^{23,25}. Owing to selection pressure, mutants that escape regulation by the synthetic population-control circuit usually arise 3 to 6 days after circuit activation (tested by dilution into fresh medium; data not shown). Although there are no direct measurements, escaping the regulation of natural systems is probably much rarer. The greater genetic stability of natural systems may be due to more sophisticated regulation or their coupling to other physiological processes, or both. For example, cell lysis is an essential stage of the development of natural transformation in *Streptococcus pneumoniae*, where survivors assimilate DNA released by lysed cells²⁵. Thus, there may be a selection pressure favouring the overall signalling network (with the lysis-regulation network as a subset) responsible for this developmental process. Design concepts such as these can be tested by introducing additional regulatory modules into the population-control circuit or coupling the circuit to functions that are beneficial to the cells.

The population-control circuit lays the foundation for using cell-cell communication to programme interactions among bacterial communities, allowing the concept of communication-regulated growth and death to be extended to engineering synthetic ecosystems. The rich repertoire of natural quorum-sensing modules^{13,14}, supplemented by engineered counterparts^{26,27}, will facilitate construction of multichannel feedback systems where population densities are coupled to communications. □

Methods

Plasmids

Plasmid pLuxR2 (ColE1 origin, chloramphenicol^R) was constructed by inserting polymerase chain reaction (PCR)-amplified *luxI* from pSND-1²⁸ into pLuxR2, downstream of *luxR*. pLuxR2 was made in two steps: (1) pLuxR encoding LuxR under the control of *Plac/ara-1* was constructed by inserting PCR-amplified *luxR* from pKE705²⁹ into pPROLar.A122 (BD Biosciences Clontech); (2) pLuxR2 was made by inserting a fragment containing *lacZα-ccdB* from pZER0-2 (Invitrogen) to *PluxI* from *pluxGFPuv*, and inserting the fused DNA into pPROLar.A122. Plasmid *pluxGFPuv* contains *p_{lux}* cloned from pKE555¹⁵. Both pLuxR2 and *pluxCcdB3* were confirmed by sequencing. Versions of the circuit constructed using different combinations of replication origins, promoters and

CcdB variants with different lethality demonstrated similar phenotypes to those reported here.

Strains, growth conditions and media

Unless otherwise stated, Top10F⁺ cells (Invitrogen) were used throughout this study. Standard LB medium was used for cell growth to probe qualitative behaviour. For quantitative measurements, cells were grown in pH-buffered TBK medium (10 g tryptone and 7 g KCl per litre buffered with 100 mM weak acids). Medium pH (measured with Accumet pH Meter 925, Fischer Scientific) was adjusted by adding 5 M KOH. MOPS was used to buffer pHs between 7 and 8, and PIPES was used for pHs between 6.2 and 6.8. The buffered media were able to maintain pH well; variations in pH were <0.3 and mostly occurred within the first 24 h of growth. Plasmids were maintained with 100 µg ml⁻¹ of chloramphenicol and 50 µg ml⁻¹ of kanamycin. One-millimolar IPTG was used to activate the circuit.

To measure circuit function, 3 ml LB in a 12 ml culture tube was inoculated from a single colony and incubated overnight at 30 °C and 250 r.p.m. When the culture reached an optical density (OD) between 0.1 and 0.3 (measured at 600 nm with a Spectra MAX250 microplate reader, Molecular Devices), the overnight culture was diluted 1,000-fold into 50 ml of fresh, buffered TBK in 250 ml flasks supplemented with antibiotics and IPTG when applicable. The flask cultures were incubated at 34 °C and 250 r.p.m. Antibiotics and the inducer were replenished about 50 h after inoculation with an additional 50 µg ml⁻¹ of chloramphenicol, 25 µg ml⁻¹ of kanamycin and 0.25 mM IPTG when applicable. Samples were drawn at different time points to measure the number of viable cells by serial dilution and plating (in triplicate), LacZ activity (see below) and OD.

LacZ assay

LacZα–CcdB levels were measured in triplicate using a FluoReporter LacZ/Galactosidase Quantitation kit (Molecular Probes). To permeate cells, 5 µl of chloroform was added to 200 µl of culture, which was vortexed vigorously for 30 s. Then 10 µl cell lysate was added to 100 µl of 1.1 mM 3-carboxyumbelliferyl-β-D-galactopyranoside (CUG) solution in each well of a 96-well microplate. The reaction mixture was incubated at room temperature for 30 min before fluorescence was measured (excitation, 390 nm; emission, 460 nm). Measured LacZ activity was corrected by subtracting the background fluorescence, measured in control wells containing 100 µl of CUG solution but no cell lysate. The corrected value was normalized to culture volume and OD 600 and expressed in fluorescence per (ml × OD 600). We observed that LacZα–CcdB caused cell death but did not destroy dead cells. Thus both live and dead cells contributed to LacZ activity. Normalizing the measured LacZ activity to total cell mass would thus give a good estimate of the intracellular killer protein concentration.

Mathematical modelling

We model the major kinetic events dictating the circuit function, including cell growth and death (equation (1)), and production and degradation of the killer protein (equation (2)) and the AHL signal (equation (3)). We assume that, (1) without the circuit, changes in viable cell density (N , ml⁻¹) follow logistic kinetics with an intrinsic per capita growth rate of k (h⁻¹) and a carrying capacity of N_m (ml⁻¹); (2) for circuit-regulated growth, the cell death rate is proportional to the intracellular concentration of the killer protein (E , nM) with a rate constant of d (nM⁻¹ h⁻¹); (3) the production rate of E is proportional to AHL concentration (A , nM, assumed to be the same inside and outside the cells) with a rate constant of k_E (h⁻¹); (4) AHL production rate is proportional to N with a rate constant of ν_A (nM ml h⁻¹); and (5) degradation of the killer protein and AHL follows first-order kinetics with rate constants of d_E (h⁻¹) and d_A (h⁻¹).

$$\frac{dN}{dt} = kN(1 - N/N_m) - dEN \quad (1)$$

$$\frac{dE}{dt} = k_E A - d_E E \quad (2)$$

$$\frac{dA}{dt} = \nu_A N - d_A A \quad (3)$$

In experiments, N and N_m are measured as colony-forming units per ml (CFU ml⁻¹). The production term ($k_E A$) in equation (2) lumps several intermediate steps together: activation of LuxR by AHL, binding of activated LuxR to p_{luxR} , and expression of the killer gene. It implies that the LuxR–AHL interaction is limited by the concentration of AHL, which is valid for our system (see main text). Although active LuxR functions as a dimer, we assume the cooperativity of AHL action to be 1, based on measurements for a closely related quorum-sensing system³⁰. Assuming a greater cooperativity does not significantly affect model predictions. Equation (3) implies that the diffusion of AHL is fast compared with other processes and that the production rate of AHL from each viable cell is the same on average.

When $N \ll N_m$, equation (1) reduces to $\frac{dN}{dt} = (k - dE)N$. Then the simplified model will have two steady-state solutions: ($N_s = 0$, $E_s = 0$, $A_s = 0$) and ($N_s = \frac{d_A d_E k}{\nu_A k_E d}$, $E_s = k/d$, $A_s = \frac{d_A d_E k}{\nu_A k_E d}$), where subscript 's' represents steady state. Linear-stability analysis shows that the trivial steady state is unstable for all positive parameters, and the non-trivial steady state is stable if and only if $d_A + d_E > k$. Because the effective degradation rate constant of the killer protein inside the cell is at least its dilution rate constant caused by cell growth, we have: $d_A + d_E > d_E > k$. Thus, the second steady state is stable for all biologically feasible parameters.

The analytical solution of the non-trivial steady state for the full model (equation (1)–equation (3)) can also be solved, in particular, $N_s = \frac{N_m d_A d_E k}{N_m \nu_A k_E d + d_A d_E k}$. This equation was used to deduce d_A (Table 1). In simulations (carried out using Dynetica³¹), the following parameters were kept at their base values: $d = 4 \times 10^{-3}$ nM⁻¹ h⁻¹, $k_E = 5$ h⁻¹,

$d_E = 2$ h⁻¹, $\nu_A = 4.8 \times 10^{-7}$ nM ml h⁻¹. The others (k , N_m and d_A) were computed from our experimental data (Table 1).

Received 18 January; accepted 15 March 2004; doi:10.1038/nature02491.

Published online 4 April 2004.

1. Becskei, A. & Serrano, L. Engineering stability in gene networks by autoregulation. *Nature* **405**, 590–593 (2000).
2. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342 (2000).
3. Atkinson, M. R., Savageau, M. A., Myers, J. T. & Ninfa, A. J. Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. *Cell* **113**, 597–607 (2003).
4. Isaacs, F. J., Hasty, J., Cantor, C. R. & Collins, J. J. Prediction and measurement of an autoregulatory genetic module. *Proc. Natl Acad. Sci. USA* **100**, 7714–7719 (2003).
5. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
6. Yokobayashi, Y., Weiss, R. & Arnold, F. H. Directed evolution of a genetic circuit. *Proc. Natl Acad. Sci. USA* **99**, 16587–16591 (2002).
7. Weiss, R., Homsy, G. E. & Knight, T. Jr *Dimacs Workshop on Evolution as Computation* 275–295 (Springer, Princeton, 1999).
8. Farmer, W. R. & Liao, J. C. Improving lycopene production in *Escherichia coli* by engineering metabolic control. *Nature Biotechnol.* **18**, 533–537 (2000).
9. Chen, W., Kallio, P. T. & Bailey, J. E. Construction and characterization of a novel cross-regulation system for regulating cloned gene expression in *Escherichia coli*. *Gene* **130**, 15–22 (1993).
10. Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nature Genet.* **31**, 69–73 (2002).
11. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
12. Blake, W. J., Kaern, M., Cantor, C. R. & Collins, J. J. Noise in eukaryotic gene expression. *Nature* **422**, 633–637 (2003).
13. Miller, M. B. & Bassler, B. L. Quorum sensing in bacteria. *Annu. Rev. Microbiol.* **55**, 165–199 (2001).
14. Fuqua, C., Parsek, M. R. & Greenberg, E. P. Regulation of gene expression by cell-to-cell communication: acyl-homoserine lactone quorum sensing. *Annu. Rev. Genet.* **35**, 439–468 (2001).
15. Eglund, K. A. & Greenberg, E. P. Quorum sensing in *Vibrio fischeri*: elements of the *luxI* promoter. *Mol. Microbiol.* **31**, 1197–1204 (1999).
16. Engelberg-Kulka, H. & Glaser, G. Addiction modules and programmed cell death and antideath in bacterial cultures. *Annu. Rev. Microbiol.* **53**, 43–70 (1999).
17. Dong, Y. H. *et al.* Quenching quorum-sensing-dependent bacterial infection by an *N*-acyl homoserine lactonase. *Nature* **411**, 813–817 (2001).
18. Leadbetter, J. R. & Greenberg, E. P. Metabolism of acyl-homoserine lactone quorum-sensing signals by *Variovorax paradoxus*. *J. Bacteriol.* **182**, 6921–6926 (2000).
19. Schaefer, A. L., Hanzelka, B. L., Parsek, M. R. & Greenberg, E. P. Detection, purification, and structural elucidation of the acylhomoserine lactone inducer of *Vibrio fischeri* luminescence and other related molecules. *Methods Enzymol.* **305**, 288–301 (2000).
20. Hasty, J., McMillen, D. & Collins, J. J. Engineered gene circuits. *Nature* **420**, 224–230 (2002).
21. Weiss, R. *et al.* Genetic circuit building blocks for cellular computation, communications, and signal processing. *Nat. Comput.* **2**, 47–84 (2003).
22. Wall, M. E., Hlavacek, W. S. & Savageau, M. A. Design of gene circuits: lessons from bacteria. *Nature Rev. Genet.* **5**, 34–42 (2004).
23. Lewis, K. Programmed death in bacteria. *Microbiol. Mol. Biol. Rev.* **64**, 503–514 (2000).
24. Ameisen, J. C. On the origin, evolution, and nature of programmed cell death: a timeline of four billion years. *Cell Death Differ.* **9**, 367–393 (2002).
25. Steinmoen, H., Knutsen, E. & Havarstein, L. S. Induction of natural competence in *Streptococcus pneumoniae* triggers lysis and DNA release from a subfraction of the cell population. *Proc. Natl Acad. Sci. USA* **99**, 7681–7686 (2002).
26. Bulter, T. *et al.* Design of artificial cell–cell communication using gene and metabolic networks. *Proc. Natl Acad. Sci. USA* **101**, 2299–2304 (2004).
27. Gerchman, Y. & Weiss, R. Teaching bacteria a new language. *Proc. Natl Acad. Sci. USA* **101**, 2221–2222 (2004).
28. Weiss, R. & Knight, T. in *Sixth International Workshop on DNA-Based Computers*, DNA 2000 (eds Codon, A. & Rozenberg, G.) 1–16 (Springer, New York, 2000).
29. Eglund, K. A. & Greenberg, E. P. Quorum sensing in *Vibrio fischeri*: analysis of the LuxR DNA binding region by alanine-scanning mutagenesis. *J. Bacteriol.* **183**, 382–386 (2001).
30. Zhu, J. & Winans, S. C. Autoinducer binding by the quorum-sensing regulator TraR increases affinity for target promoters *in vitro* and decreases TraR turnover rates in whole cells. *Proc. Natl Acad. Sci. USA* **96**, 4832–4837 (1999).
31. You, L., Hoonlor, A. & Yin, J. Modeling biological systems using Dynetica—a simulator of dynamic networks. *Bioinformatics* **19**, 435–436 (2003).
32. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1–12 regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210 (1997).

Acknowledgements Y. Wang, R. Georgescu, S. Thiberge, F. Balagadde and S. Maerle assisted with preliminary characterization of the circuit. C. Collins constructed plasmids pLuxR, pLuxR2 and pLuxGFPuv. We also thank Y. Yokobayashi, M. Raizada, J. Leadbetter, M. Elowitz and M. Savageau for discussions or comments on the manuscript. This material is based on work supported by the Defense Advanced Research Projects Agency (DARPA). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the DARPA.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.H.A. (frances@chem.caltch.edu).

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs Sales Director:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Scandinavia/ Spain/ Portugal:

Evelina Rubio Håkansson (4973)

Natureevents: Silke Opstrup (4994)

France/ Switzerland:

Amelie Pequignot (4974)

Production Manager:

Billie Franklin
To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Stefan Hales

European Satellite Office

Germany/ Austria/ Italy/

The Netherlands/ Belgium:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatamachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami

e-mail: rasami@naturejpn.com

naturejobs

Thinking big

Big science means big opportunities — huge centres, multimillion-dollar contracts and, often, hundreds of funded positions for investigators, postdocs and students. But choosing to take part in such enterprises, at any level, also comes with big risks — or at least some pretty serious questions about job security.

Take the announcement earlier this year that the US government plans to abandon the Hubble Space Telescope (see *Nature* **427**, 574; 2004). Hundreds of scientists are involved in ground support at multiple centres in the United States and Europe. When Hubble stops taking data, perhaps as soon as 2007, will the support staff stay on with the same institutions? Or will they slowly be phased out by attrition of contracts?

The answers remain unclear. But Steve Beckwith, director of the Space Telescope Science Center in Baltimore, Maryland, which provides ground support for Hubble, says that as things stand, the end of the project could mean layoffs for about half the centre's workforce over the next two to three years. The lack of clarity comes from the possibility that the institute's staff may still be of service — there are proposals to maintain Hubble using robotics. But such an endeavour has never been successfully done, so it is unclear whether it will be funded, planned and staffed in time.

Similar questions can be asked of job security at other centres. What will happen to scientists at Fermilab in Batavia, Illinois, when the larger, more powerful Large Hadron Collider makes its equipment effectively obsolete?

The moral of Hubble's recent news is to evaluate the longevity of a big project before you sign on. And try to develop skills that will be relevant to another effort once that big project inevitably comes to an end.

Paul Smaglik
Naturejobs editor



Contents

REGIONS

Benelux goes for growth p874

CAREER VIEW

Recruiters & Academia

E-recruiting

Graduate Journal

A foot in two doors

Movers

Matthew Sherman

p876

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

Fertile ground Benelux



Whenever Marc Van Montagu looks out of his office window, the plant-science pioneer sees a reminder of a field he seeded in Belgium. Twenty years ago, he and the late Josef Schell founded Plant Genetic Systems (PGS), one of the first companies to exploit the potential of gene transfer in plants. The technology went on to revolutionize plant breeding.

Through a series of mergers and acquisitions, PGS became Bayer CropScience — and next month more than 180 scientists and support staff will move into labs in a new building opposite Van Montagu's office in the Ghent Technology Park.

"I am really happy it stayed in Ghent," Van Montagu says of the company he helped found. He is well aware that plant science is itself a delicate flower, threatened by declining funds and lack of public acceptance throughout the European Union (EU), whose citizens are wary of genetically modified organisms (GMOs).

Public opinion is no warmer towards GMOs in Belgium, the Netherlands and Luxembourg — known collectively as 'Benelux' — than anywhere else on the continent. However, these countries have built up unique strengths not only in genetic engineering but also in traditional plant breeding, nutrition, food safety and environmental research.

As a result, this culturally diverse and densely populated region comprises the continent's most

fertile grounds for food and plant science. "There is no such concentration of agricultural biotechnology elsewhere in Europe," says Michiel Van Lookeren Campagne, head of research at Bayer CropScience.

The Dutch-speaking Belgian province of Flanders typifies Benelux's fruitful mix of strong academic research, entrepreneurial spin-off and research-intensive industry. Similar academic-industrial clusters exist around Liège and Namur in Wallonia, Belgium's French-speaking province. The 'Food Valley' in the Netherlands, centred around Wageningen, provides plenty of space, support and opportunities for all kinds of activities in agricultural biotechnology. In Luxembourg, the Gabriel Lippmann Research Centre has a department of environmental and biotechnology research that covers a range of activities from forestry to agricultural science.

ENGLISH SPOKEN

The Benelux countries don't burden individual researchers with excessive administration. Foreigners find it easy to adapt — not only in the lab, but also in everyday life, as English is widely spoken.

"In my last job, with Bayer in Germany, I had huge language difficulties, not only in public but also with technicians," says Australian postdoc Ryan Whitford, who last year joined the Flanders Interuniversity

VIB

Institute for Biotechnology (VIB). "Here it is much better — and the beer is fantastic." Whitford, a molecular plant biologist, was attracted to both industry and academia. He opted for the VIB, where he has the best of both worlds. There he works with model organisms, a staple of basic research, but always with an eye on industry applications.

The VIB, a virtual institute founded in 1995, supports 60 independent medical and plant research groups involved in the life sciences in Ghent, Leuven, Brussels and Antwerp. They use a mix of molecular and cell biology, genetics and bioinformatics to speed the translation of basic research into practical applications in medicine, plant breeding and agriculture.

In the Ghent Technology Park, the VIB's plant scientists share a brand-new lab building with the department of molecular biomedical research. Small

thousands of modified rice plants are automatically handled and digitally photographed at different growth stages. As Wim Van Camp, the company's director of technology management, says, it's "an enormous challenge in terms of data management".

Life scientists in Belgium and the Netherlands find it easy to switch between academic and industrial research. Postdocs in industry, such as Philippe Hervé, a French plant biologist with CropDesign, can become team leaders quite early in their careers. But they then spend less time at the bench, and more in meetings, than colleagues at universities and research institutes.

In the Netherlands, the food industry accounts for 10% of the gross national product, 10% of employment and more than a third of exports. The Wageningen Centre for Food Sciences (WCFS) fosters public-private partnerships in the Food Valley, Europe's largest centre of food-related research. Set up in 1997 with help from the Dutch government, it promotes cooperation in pre-competitive research among seven large food companies, including Dutch-British giant Unilever.

A TASTE OF SUCCESS

The WCFS supports research related to food and health in areas from nutrigenomics — how nutrients interact with the genome — to food structure, safety and microbial functionality. It is linked with graduate schools at the universities of Wageningen and Maastricht, which offer additional courses to 50 or so PhD students at a time. Each year, some 30 post-doctoral positions become available, in areas from pure science to product development.

"One interesting new field is the molecular basis of sensory properties," says Martijn Katan, scientific director of the WCFS nutrition and health programme. "What is it that makes something taste nice or creamy in your mouth?" Each company has only a minor influence on the research programmes, says Katan. Indeed, he adds, the WCFS is involved in evaluating industry claims about the health benefits of research-intensive 'premium products', such as probiotic yoghurts containing allegedly beneficial bacteria and spreads containing plant substances to reduce cholesterol uptake.

Wageningen is an ideal environment for agricultural biotechnology, just as Leiden, 100 kilometres west, is for the medical life sciences. Labs and small enterprises are spread all over these towns, easily reached by bicycle. Students, scientists and employers encounter each other at every corner and in every pub.

Aalt Dijkhuizen, president of the Wageningen University and Research Centre, prides himself on the tightly focused research at the small university, where people from 120 nationalities work in the whole range of plant, nutrition, animal and environmental sciences.

Their skills are in demand locally. Keygene, for example, is a supplier of molecular markers and genetic 'fingerprinting' to aid conventional plant breeding. Perhaps because it offers an alternative to GMOs, the company has grown from four employees in 1989 to more than 100 now, with further plans for expansion.

Van Montagu is optimistic that applying such high-tech approaches to conventional breeding will help to solve GMO problems — and ensure that the plant-science tradition he planted will continue to flourish. ■

Quirin Schiermeier is *Nature's* German correspondent.

Q. SCHIERMEIER



Fields of endeavour: plant science bears fruit in Flanders, opposite, and at CropDesign, where Wim Van Camp, above left, works with his postdocs.

groups gather around cube-shaped wooden tables in the spacious lobby, to the delight of Jo Bury, the VIB's director general. "Communication is everything," he says, "and you can bet they're talking about science."

The VIB's department of plant genetics was set up in 1996 under Van Montagu's leadership, and renamed the department of plant systems biology after he retired in 1999. Its focus now is on studying and modelling cell division and cell-cycle control in plants.

This redefinition has created the need for people with skills beyond the plant-science community. "You need very diverse disciplines to do this," says Martin Kuiper, head of the institute's computational biology division. "Unfortunately, there are not many biologists around with really good informatics skills."

Bioinformaticians are also in demand at two VIB spin-offs on the same campus. The first, Devgen, was founded in 1997 by Thierry Bogaert, a professor of biochemistry at the University of Ghent, to identify new drug targets in the nematode *Caenorhabditis elegans*. The company also develops chemical and genetic solutions for crop protection.

A stone's throw away at CropDesign, researchers are using high-throughput technologies to validate the effect of gene modification on the growth and yield of cereal crops. In the high-security greenhouses, tens of

Flanders Interuniversity
Institute for
Biotechnology

♦ www.vib.be/VIB/EN

Wageningen Centre for
Food Sciences

♦ www.wcfs.nl

Wageningen University
and Research Centre

♦ www.wur.nl/uk/index.html

Centre for BioSystems
Genomics

♦ www.cbcs.nl

BioPartner, for
companies starting up
in the life sciences

♦ www.biopartner.nl/asp/default.asp

CropDesign

♦ www.cropdesign.com

Devgen

♦ www.devgen.com/index.php

Keygene

♦ www.keygene.com/index2.htm

Unilever

♦ research.unilever.com

Check-Points, maker of
a microbial test for food

♦ www.check-points.com

GRADUATE JOURNAL

A foot in two doors

When I decided to leave my industry job to earn a master's degree, I knew I would return. Consequently, I sniffed out opportunities, went to interviews and even got a few job offers before I handed over my lab keys. When I left industry for the second time to do a PhD, I again assumed I would be back. But this time I did not feel the same drive to have a safe industry offer.

While talking to a senior scientist at a conference (see *Nature* 427, 570; 2004), he pointed out that great research careers can spring from working on the leading edge of a new field. If given the option, he opined, a young researcher should run for risky, less-mature fields of research, and work to be part of a new wave.

Of course, that is easier to say when one holds a secure senior research position, but I think he has a point. Hunting down a postdoc position in a challenging and emerging field may be the least risky career move I make.

The job market will always fluctuate, and no one knows which research fields will boom in the future. There will be few times when I can follow my instinct and see what happens. Still, I won't make that jump without sending change-of-address notes to my industry contacts. ■

Sidney Omelon is a PhD student in bone biomaterials at the Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada.

RECRUITERS & ACADEMIA

E-recruiting

In academia, researchers traditionally use osmosis to build multidisciplinary collaborations — scientists diffuse into a new field by reading research in an area, then picking their partners. But the splitting of science into smaller units and the explosion in the amount of published literature has made creating such connections difficult.

This problem can be overcome by using the Internet. Hosting profiles of users can serve as a worldwide magnet for like-minded, if far-flung, kindred spirits, and can simplify the process of bringing scientists together.

Over the past five years I have used such e-recruiting techniques to help seed the growth of a new field called 'telepreventive medicine', which uses the Internet to educate people about preventive medicine. The network has grown from fewer than 50 individuals

in 1999 to more than 13,400 scientists from 151 countries this year. The mission is to tap into every scientist interested in disease prevention, and to provide, for free, a 'supercourse' — an open-source set of PowerPoint lectures on prevention for educational usage worldwide.

The e-recruitment strategy combines technology with social currency to boost involvement. For example, scientists working in epidemiology naturally want to help arrest the spread of disease. So they share lectures because they know their work will benefit others — especially colleagues in the developing world.

The network has expanded using both low- and high-tech methods. On the low-tech side, many on the network tell their friends and faculty members who, in turn, become users. And, as a

group, we annually distribute 10,000 CDs with 1,000 lectures on them and ask recipients to pass copies to colleagues and students.

On the high-tech side, the network uses weekly updates to alert users to additions. It can also respond quickly to a crisis. For example, should a tornado hit Oklahoma, an announcement for information can be sent out through the network and within hours there will be a 'tornado supercourse' set of experts, accompanied by lectures that people can tap into.

The next step may be to widen the network beyond public health into all of science. I envisage a broader application that could bring together scientists with common interests but from different disciplines. ■

Ronald LaPorte is a professor of epidemiology at the University of Pittsburgh, Pennsylvania, USA.
♦ www.pitt.edu/~super1

MOVERS Matthew Sherman, senior vice-president, Synta Pharmaceuticals, Lexington, Massachusetts



Matthew Sherman's career path seems almost cyclical. Emerging from an academic background in Boston and Cambridge, Massachusetts, he went to work for the Genetics Institute, a small biotechnology company in Cambridge. That was eventually acquired by the drug giant American Home Products based in Pennsylvania. Now, some 12 years later, Sherman finds himself back in the Boston area as senior vice-president and chief medical officer for another small company, Synta Pharmaceuticals.

Sherman attributes his career direction to Alexander Rich, a professor at the Massachusetts Institute of Technology with whom Sherman worked as an undergraduate. Rich was an MD who had opted for a career in X-ray crystallography rather than clinical practice. He pointed out to Sherman that the traditional path from MD into practice or PhD to tenure was not obligatory.

Sherman took that advice and after earning his MD began a fellowship that allowed him to split his time between patient care and oncology research. That led to a position where he could add teaching to his research and clinical practice. "I love all three," Sherman says.

During his nine years in academia, Sherman began cultivating an entrepreneurial side. "I was interested in combining all the skills I'd learned in the lab with the business skills of what it took to put a company together and to develop a product," Sherman says.

So in 1992, he bit the bullet and left academia to join the Genetics Institute. There he was able to learn the business of drug discovery and development. But after five years of on-the-job training, the company was bought by American Home Products and Sherman found himself working for a very large firm.

Now that he has spent time in such an environment he is ready for a new challenge. "I wanted to take the skill sets I learned and apply them to the small biotech setting at Synta," he says.

The Synta job appealed to him because, having specialized in oncology and immunology, he can now work across all disease groups and have a broader involvement with the drug-development process. And in doing so, he wants to keep the spirit of good mentoring alive. "As a professor, your biggest impact is on the students you teach. In industry, it's the employees under you," says Sherman. ■

CV **1992–2004:** Associate director, clinical research rising to clinical site head, clinical research and development, Wyeth-Ayerst Research/Genetics Institute, Cambridge, Massachusetts and Collegeville, Pennsylvania
1983–1992: Fellow in medical oncology rising to assistant clinical professor of medicine, Dana-Farber Cancer Institute, Boston, Massachusetts